SHMcloud™ User Manual

eDiscovery processing, on your workstation or in the cloud

User Manual updated: 04/25/13

SHMsoft, Inc.
One Riverway
Houston, Texas 77056
info@shmsoft.com
tel: (713) 568-9753
fax: (206) 339-8596
http://shmsoft.com

Table of Contents

Introduction
About SHMcloud
Minimum System Requirements
SHMcloud™ eDiscovery processing on Hadoop clusters using Amazon EC2 instances
Summary
Installation
What happens if you try to run “shmcloud_player” before you have extracted your files?
Moving Forward
2. Run the SHMcloud™ Player by double-clicking on “shmcloud_player”.
Getting Started - Testing SHMcloud
6. Processing your test job
What is staging?
7. Process Locally
8. Reviewing the results
. Report
9. Metadata
9.3 Standard metadata fields.
. Native Zip Folder
Exception Folder
Native Folder
Text Folder
10. Creating & saving your own project
Points to notice:
10.15 Now we are ready to Process our project.
Notes and Warnings:
⇛ metadata - is your project output load file, as discussed in detail in Section 9. This is the output that you are looking for when you run your project.
⇛ native - is a zipped folder. It contains all extracted native files, including emails and text extracted from them, as well as “exception” files that could not get processed for any reason. Essentially it is everything that this project processed.
⇛ report - is a simple report of your run. It contains the name of your project, when it started, when it finished, how long it took to run, and how many items were included in this run.
Multiple Output Files
11. Setting up an Amazon AWS Account
12. Processing your project in the Cloud
Bucket & Project Notes:
S3 - Abridged steps
Moving Forward:

13. Amazon's Strong Security on EC2
Setting up a Security Group
Setting up Key Pairs
Preparing your EC2 (Elastic Compute Cloud) for processing
14. Cluster Control - How to Turn on your Cloud Computer & Run Your Project on Amazon
14.4 Shutting Down the Cluster
How can you determine that the cluster really turned off?
14.5 - Reviewing your output after running your project on Amazon
15. Creating Projects With Specialized Searches
Installing Solr on your computer for use with SHMcloud™
How to run projects in conjunction with the Solr Search Server:
Turning on your Solr Search Server
Step 1 - Stage your project.
Step 2 - Process locally, and wait for your project to complete processing.
Step 3 - Review your output in Solr.
Error Messages while attempting to access Solr
Viewing all of your processed documents at one time:
Searching all of your data manually, while using a standard search function:
Refining your search, using the Solr Search features:
Additional Search Options in Apache-Solr:
Custodian Search:
Date Range Search:
Test String Search:
Author Search:
Combining Search Options:
Licensing

Introduction

This software is intended for use by lawyers, litigation support specialists, compliance and forensics analysts, pro-se litigants, and in general for custom searches in files.

This software does eDiscovery processing: text extraction, culling, and native/text/metadata delivery. It consists of the desktop application, called SHMcloud™ Player and the SHMcloud itself, the processing backend on Amazon AWS computers.

You can use the Player for local processing, if you computer is powerful enough, and if the amount of time it will take on one machine is acceptable. This processing is free. If you want to use the cloud, you upload the files using the Player, and direct the SHMcloud to do the processing. In this case, AWS machine charges will apply.

SHMcloud versions and capabiliies.

Capability	**Standalone Player in Windows**	**Standalone Player in Linux**	**EC2 processing (no setup needed)**
OCR	No	Yes, but a setup is required	Yes
Imaging	PDF	PDF	PDF, in testing
Search	Yes, Solr setup required (simple)	Yes, Solr setup required (simple)	In the works, coming soon

About SHMcloud

Thank you for choosing SHMcloud™.

SHMsoft is a Big Data applications solutions provider. The company was first in pioneering the concept of Hadoop-based e-discovery to serve Global 2000 companies confronted with the task of managing highly complex, heterogeneous and decentralized IT environments in a world that is constantly and rapidly changing.

Users are encouraged to email any questions and feature requests to info@shmsoft.com.

SHMcloud™ is a complete large-scale data processing, search and analytics solution for e-discovery utilizing the latest Hadoop/MapReduce/HBase technologies. Hadoop allows you to put terabytes of data in one place. But more than just a container, SHMsoft's Hadoop distribution allows you to regain control over your data by allowing you to process, analyze and review your own data in-house during litigation or for any business requirement.

If you have 100 Gigs to process, you can spin up 50 machines on AWS EC2 with our Hadoop clusters and have the work finished in about an hour. See how we did it with Enron data here, http://shmsoft.blogspot.com/2012/06/processing-enron-data-on-49-node.html.

SHMcloud™ processes large data sets across clusters of computers that are designed to scale up from single servers to hundreds of machines, each offering local computation and storage. Processing is organized by the Hadoop framework. Each file is read from the archive, assigned a unique ID, and processed with Tika, which extracts text and metadata. Metadata, text, and the file itself are delivered as processed results.

With this compilation and professional support available for enterprise use, SHMcloud brings high performance, scalability and reliability to data processing at a fraction of the cost of proprietary products.

Suggested Minimum System Requirements

For the SHMcloud™ player

2 GB of RAM
5 GB Hard Drive Space
Java 7.0 and higher
Supported Operating Systems include:
Windows XP, Windows 7 and Vista
Linux
Mac OS X

Nota bene: If you want to use your SHMcloud player for local processing, then use as powerful a workstation as possible.

For the SHMcloud™

Internet speed should be fast. There are upload and download operations, and you don't want them to go for too long.

Machines used in the cloud are currently hard-coded, but later there will be a choice. However, even now you can find that parameter in the setting properties file in the install directory. The two choices are c1.medium and c1.xlarge.

Number of nodes in the cluster is currently recommended to be set from 5 to 10. Later, when we implement parallel operations on startup, this number will be increased.

Recommended size for staging archive is between 1 GB to 5 GB.

Please Note: If you do not have Java properly installed on your system, then your SHMcloud™ Player will not run.

Java, can be downloaded for free from oracle.com. If you have difficulty with setting the proper path parameters for your Java install, then please contact SHMsoft at http://shmsoft.com/, and we will be happy to assist you.

SHMcloud™ eDiscovery processing on Hadoop clusters using Amazon EC2 instances

The next few pages will include more detailed instructions for running SHMcloud™.

Summary

⇨ Open the SHMcloud™ Player on your computer. Do this by double-clicking on “run_gui” in the SHMcloud™ folder (“run_gui.bat” for Windows, and “run_gui.sh” for a unix-based environment).

⇨ Several windows will open, including the main application window, a Processing History window, and a command window. The main window has SHMcloud™ in the title.

⇨ First, you will create a new Project to be processed in SHMcloud™. You will define the project and the files to be used, and you will stage the data.

⇨ Next, you will setup access to your Amazon environment, including S3 and EC2.

⇨ And finally, you will process your project, which entails uploading content to Amazon, processing it, and downloading results from Amazon. ***Fortunately, the SHMcloud™ application performs these tasks for you, making the entire process quite easy.***

Now let’s get started.

Installation

1. Download and install the SHMcloud™ player by unzipping it to an easy to find location. You will need a code key to perform the unzip action which you will get very quickly by sending an email to FreeEed-key@shmsoft.com and requesting a copy of the key. Please provide your name, company name, and telephone number in your email.

Once your Download is complete you will need to unzip, or extract, the files.

(1) Right-click on your zipped SHMcloud folder.
(2) Select “Extract All...” from the menu.
(3) A Destination will be suggested. Is this where you want your SHMcloud folder to go? If yes, then click on the Extract button.
(4) You will be told that a Password is required. Did you send an email requesting the code, as explained above? Enter the code that you received in your email from FreeEed-key@shmsoft.com.

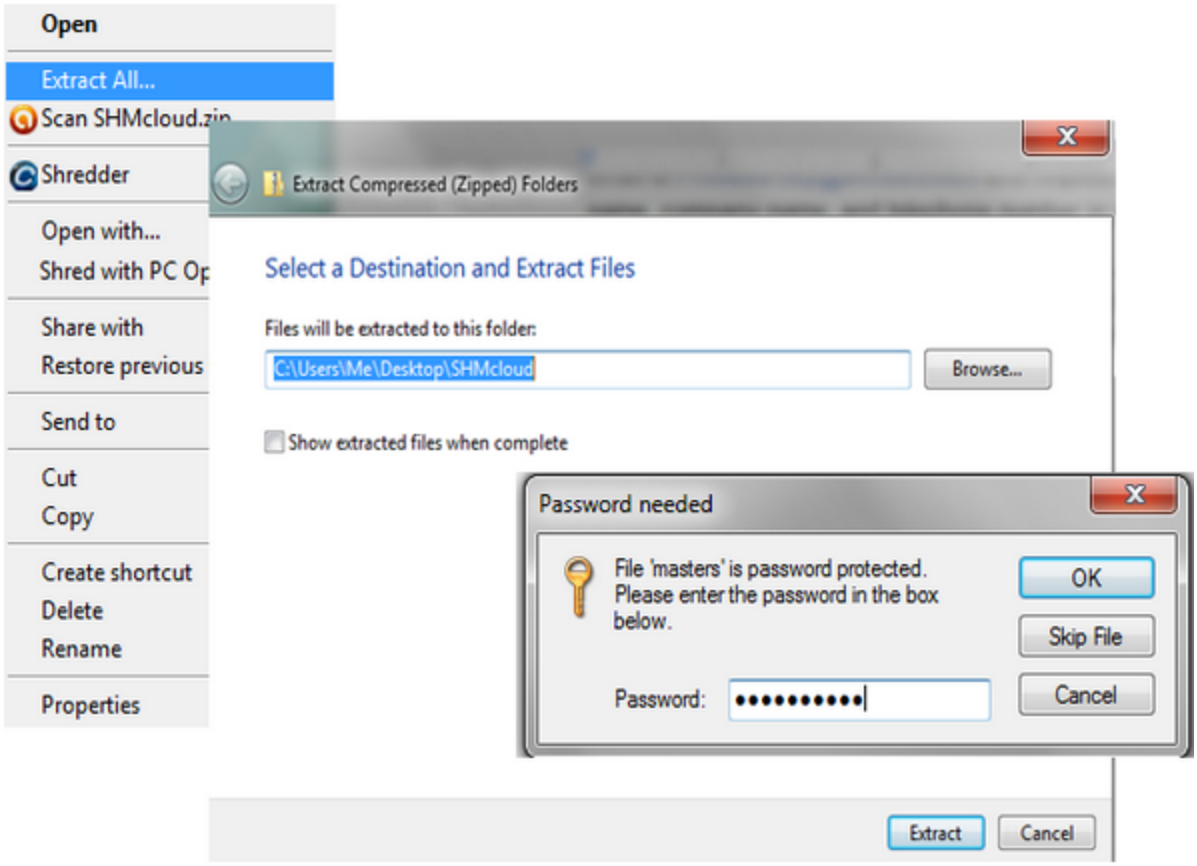


FIGURE 1.0

Figure 1.1 shows you what you should have inside your SHMcloud™ Folde after the file is unzipped.

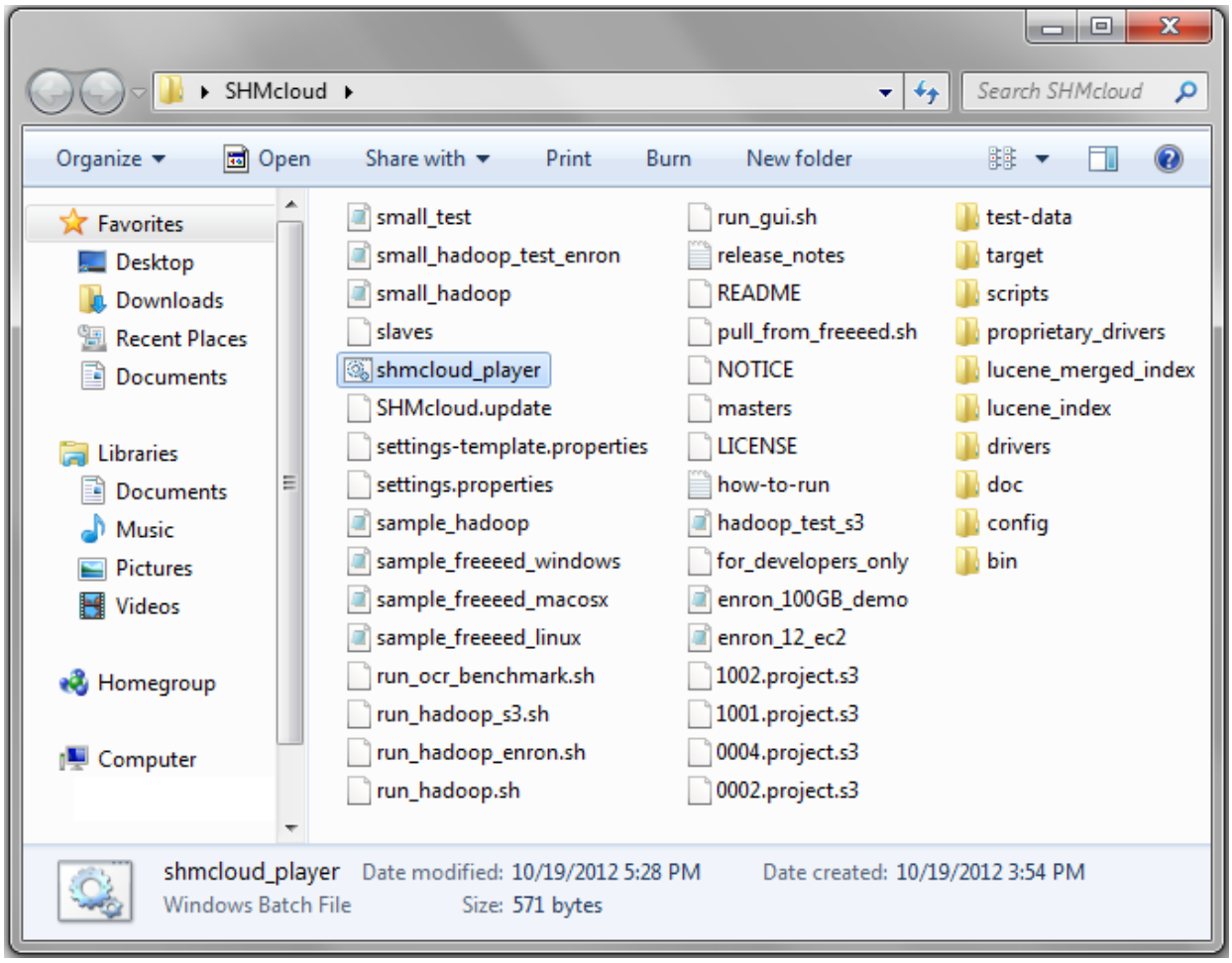


FIGURE 1.1

In #2, we will double click on shmcloud_player to run the SHMcloud™ Player. But what will happen if you did not extract your folder by following the steps above? What will happen if you double clicked on your zipped folder and found shmcloud_player and decided to run it from your zipped folder?

What happens if you try to run “shmcloud_player” before you have extracted your files?

If you are inside your zipped folder then double clicking on shmcloud_player will bring up a small screen, Figure 1.2. It is necessary for you to select “Extract all” in order to unzip the file.

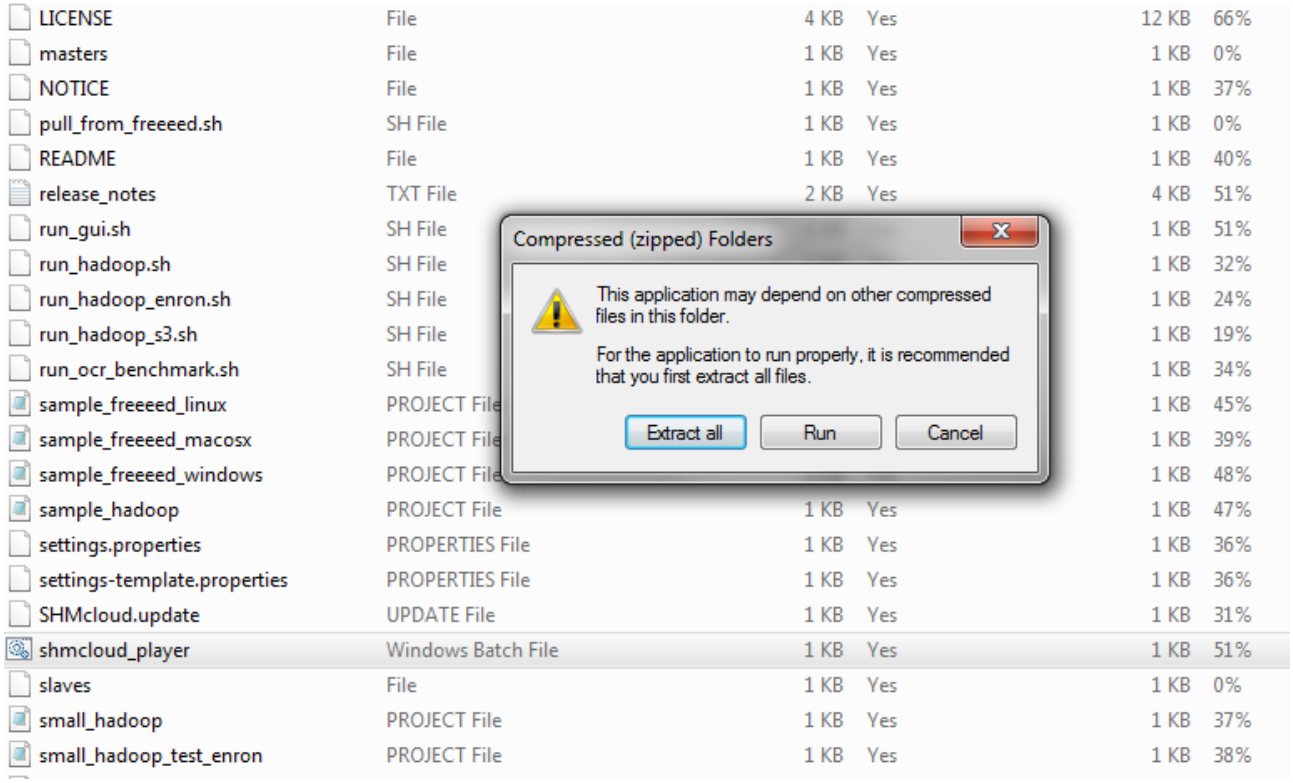

LICENSE File 4 KB Yes 12 KB 66%
masters File 1 KB Yes 1 KB 0%
NOTICE File 1 KB Yes 1 KB 37%
pull_from_freeeed.sh SH File 1 KB Yes 1 KB 0%
README File 1 KB Yes 1 KB 40%
release_notes TXT File 2 KB Yes 4 KB 51%
run_gui.sh SH File 1 KB 51%
run_hadoop.sh SH File 1 KB 32%
run_hadoop_enron.sh SH File 1 KB 24%
run_hadoop_s3.sh SH File 1 KB 19%
run_ocr_benchmark.sh SH File 1 KB 34%
sample_freeeed_linux PROJECT File 1 KB 45%
sample_freeeed_macosx PROJECT File 1 KB 39%
sample_freeeed_windows PROJECT File 1 KB 48%
sample_hadoop PROJECT File 1 KB Yes 1 KB 47%
settings.properties PROPERTIES File 1 KB Yes 1 KB 36%
settings-template.properties PROPERTIES File 1 KB Yes 1 KB 36%
SHMcloud.update UPDATE File 1 KB Yes 1 KB 31%
shmcloud_player Windows Batch File 1 KB Yes 1 KB 51%
slaves File 1 KB Yes 1 KB 0%
small_hadoop PROJECT File 1 KB Yes 1 KB 37%
small_hadoop_test_enron PROJECT File 1 KB Yes 1 KB 38%
Compressed (zipped) Folders
This application may depend on other compressed files in this folder.
For the application to run properly, it is recommended that you first extract all files.
Extract all
Run
Cancel

FIGURE 1.2

When you select “Extract all”, two windows will pop up. The first window (Figure 1.4) might jump behind the SHMcloud window, and if you blink you might miss it. Can you see the shadow in the background behind the SHMcloud screen in Figure 1.3? As in Figure 1.3, the second window will remain on the top of your screen. This is where you should enter the code that you received after sending an email to FreeEed-key@shmsoft.com.

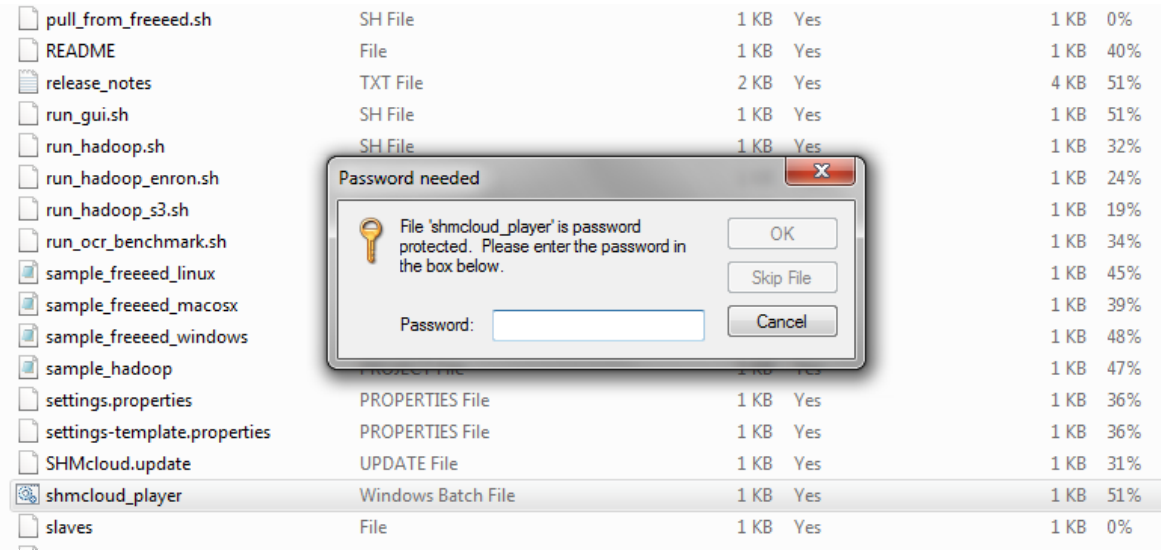

pull_from_freeeed.sh SH File 1 KB Yes 1 KB 0%
README File 1 KB Yes 1 KB 40%
release_notes TXT File 2 KB Yes 4 KB 51%
run_gui.sh SH File 1 KB Yes 1 KB 51%
run_hadoop.sh SH File 1 KB Yes 1 KB 32%
run_hadoop_enron.sh 1 KB 24%
run_hadoop_s3.sh 1 KB 19%
run_ocr_benchmark.sh 1 KB 34%
sample_freeeed_linux 1 KB 45%
sample_freeeed_macosx 1 KB 39%
sample_freeeed_windows 1 KB 48%
sample_hadoop 1 KB 47%
settings.properties PROPERTIES File 1 KB Yes 1 KB 36%
settings-template.properties PROPERTIES File 1 KB Yes 1 KB 36%
SHMcloud.update UPDATE File 1 KB Yes 1 KB 31%
shmcloud_player Windows Batch File 1 KB Yes 1 KB 51%
slaves File 1 KB Yes 1 KB 0%
Password needed
File 'shmcloud_player' is password protected. Please enter the password in the box below.
OK
Skip File
Password:
Cancel

FIGURE 1.3

After you enter the key into the Password box, that window will close. If you do not see the Extract window, look for it behind your SHMcloud files screen, then select Extract.

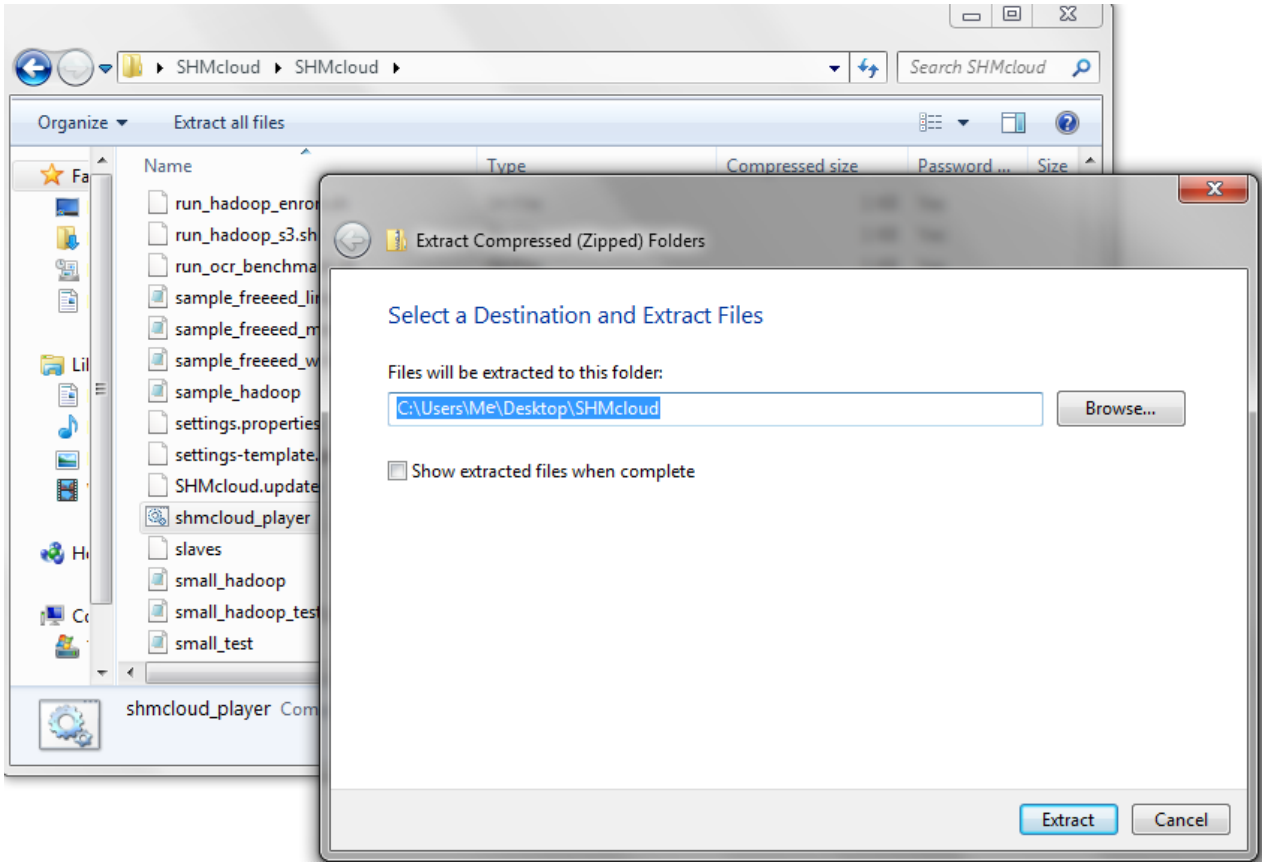


FIGURE 1.4

Once the file is extracted, an unzipped SHMcloud folder will appear in the area that is designated

in Figure 1.4.

Open the unzipped folder. If there is another SHMcloud folder in there, then open that one until you see the folder contents.

If you did not follow the Extract instructions this is listed in #1, then you may still have files to extract from your zipped folder. So while you are working you may be asked to enter the key a second time. Entering the password again should extract the rest of your files. Make sure to move to the new SHMcloud directory that the program created for you, or it will keep asking for your password every time you try to move forward.

We recommend that you do not extract your files by clicking into the zipped folder, but rather you should *right-click on the zipped folder* as explained above at the start of #1.

Troubleshooting: What if you go away, come back at some point in the future, restart your Player by clicking on **shmcloud_player** and suddenly you are asked to provide the Password, but you know you already extracted the files? Check again, you probably clicked into your zipped up SHMcloud™ folder! Try again and look for the SHMcloud™ folder without the zipper on it!

Note: At this point, if you prefer, you can create a shortcut to your shmcloud_player onto your desktop for easy access. If you choose to do this, make certain to do so by using the “create shortcut” feature provided by your computer. Simply copying and pasting the shmloud_player onto your desktop will not work.

Moving Forward

Before you can run the next step you need to have the most recent version of Java installed on your computer. If you do not have Java, then you can download a free version by going to oracle.com and selecting the Free Java Download.

2. Run the SHMcloud™ Player by double-clicking on “shmcloud_player”.

You will activate three screens which may be tiled on top of each other. What are these three screens? The SHMcloud™ window is your action window. This is your SHMcloud™ Player, used to process your Projects. The “History” screen and the CMD screen will be running in the background during processing. These screens will give you useful information about your processing job. When your SHMcloud™ Player has completed processing, you will see the word, “Done” displayed at the bottom of the History Screen.

Note: If “shmcloud_player” will not run, then in all likelihood Java is not installed properly on your computer. Go to the command (DOS) window, type the word “JAVA”, and hit enter. If Java is not installed or is not recognized in the command line, then please contact SHMsoft at http://shmsoft.com and we will help you to reset your path parameters.

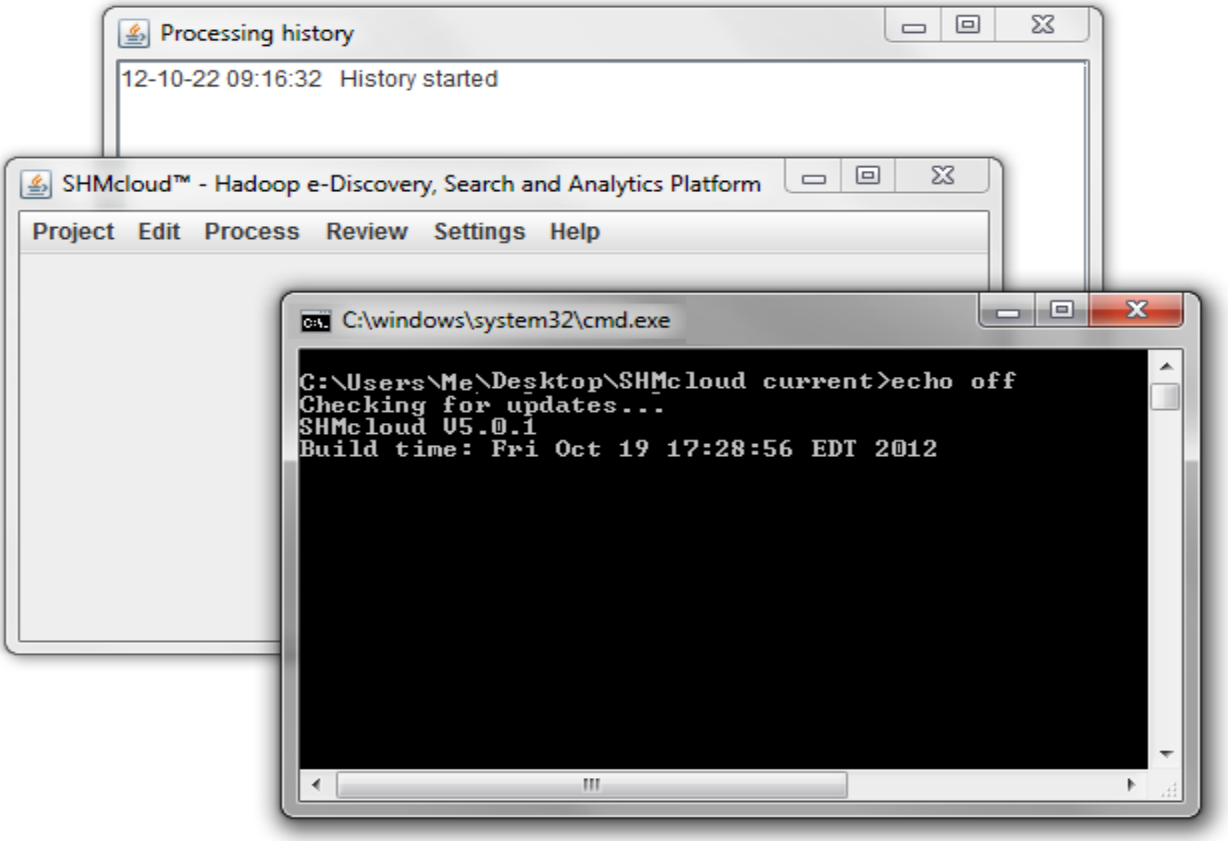

Processing history
12-10-22 09:16:32 History started
SHMcloud™ - Hadoop e-Discovery, Search and Analytics Platform
Project Edit Process Review Settings Help
C:\windows\system32\cmd.exe
C:\Users\Me\Desktop\SHMcloud current>echo off
Checking for updates...
SHMcloud V5.0.1
Build time: Fri Oct 19 17:28:56 EDT 2012

FIGURE 2

Once your Player opens these three windows, you may close your SHMcloud™ folder, seen in Figure 1.1. Henceforth your files will be accessed directly from within the SHMcloud™ Player.

Getting Started - Testing SHMcloud

3. There is a test job supplied with SHMcloud™ that you can run in order to verify that all the files transferred correctly to your machine, and to verify that your platform meets the minimum requirements. To perform the test supplied with the program, pull down the “Project” Menu and select “Open”.

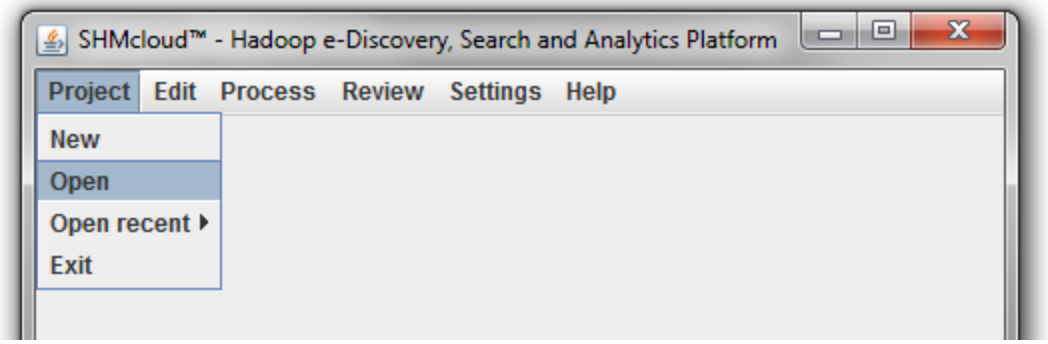


FIGURE 3

We recommend that you run through the test project just to make sure everything is working properly. In section #10 we will begin to show you how to process your own projects.

4. The Open command will bring up a window that looks like Figure 4 below. Select the project “sample_freeeed_windows.project” by double clicking on it.

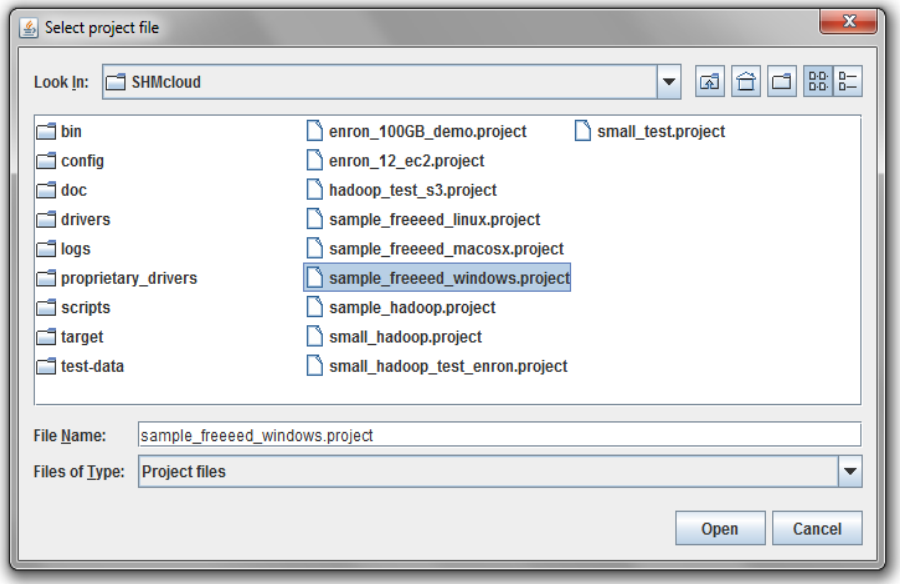
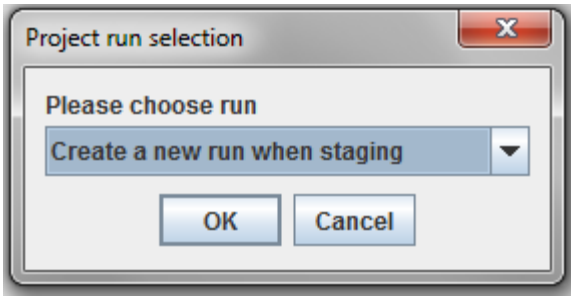


FIGURE 4

Since this particular project already exists in your SHMcloud™ Player, you will be asked to choose a run, or create a new run when staging. Because you have not yet run this project on your own computer, you will have to select “create a new run when staging”.



5. After you double click on the file “sample_freeeed_windows.project” a window with the project settings opens (if this is a new clean project).

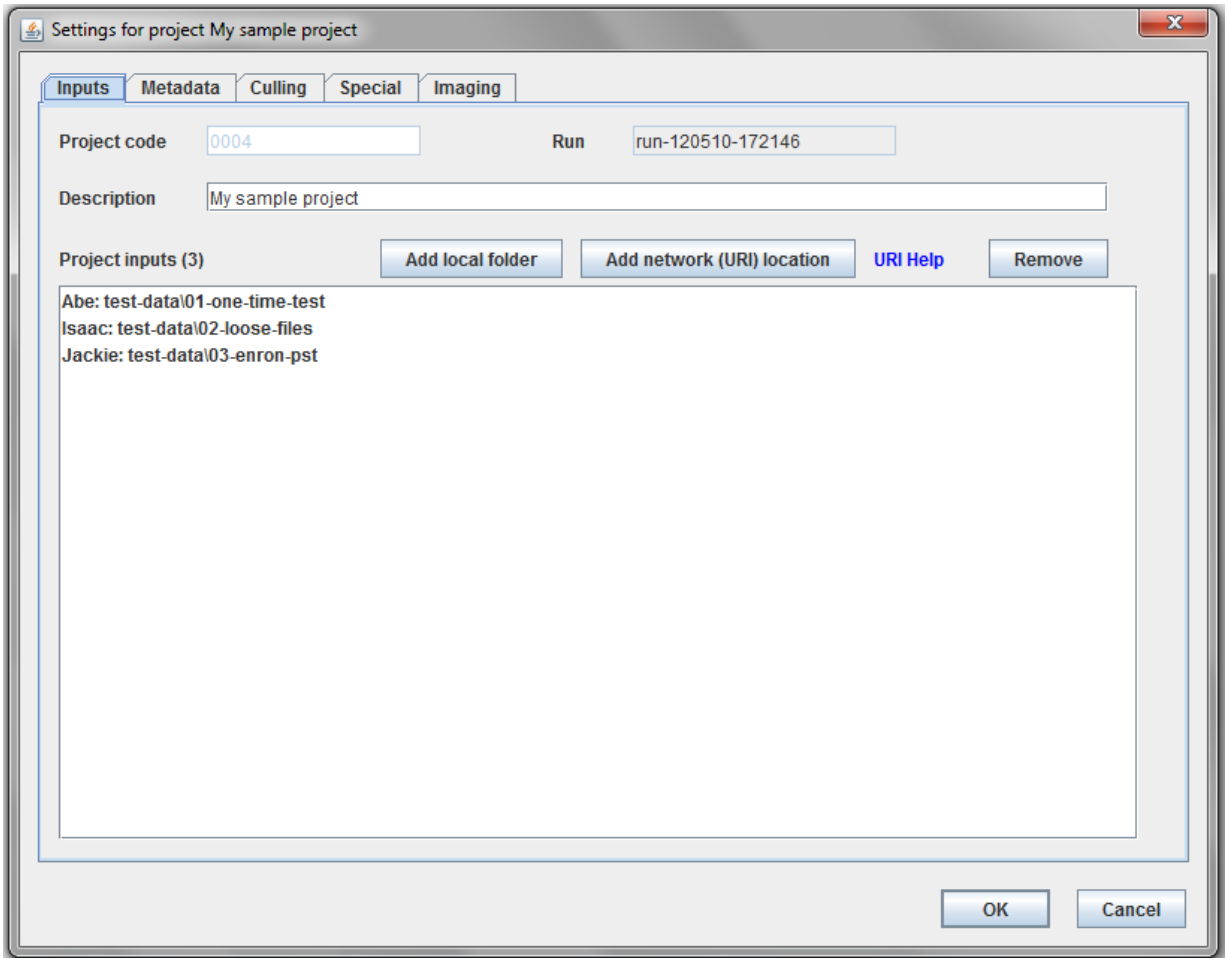


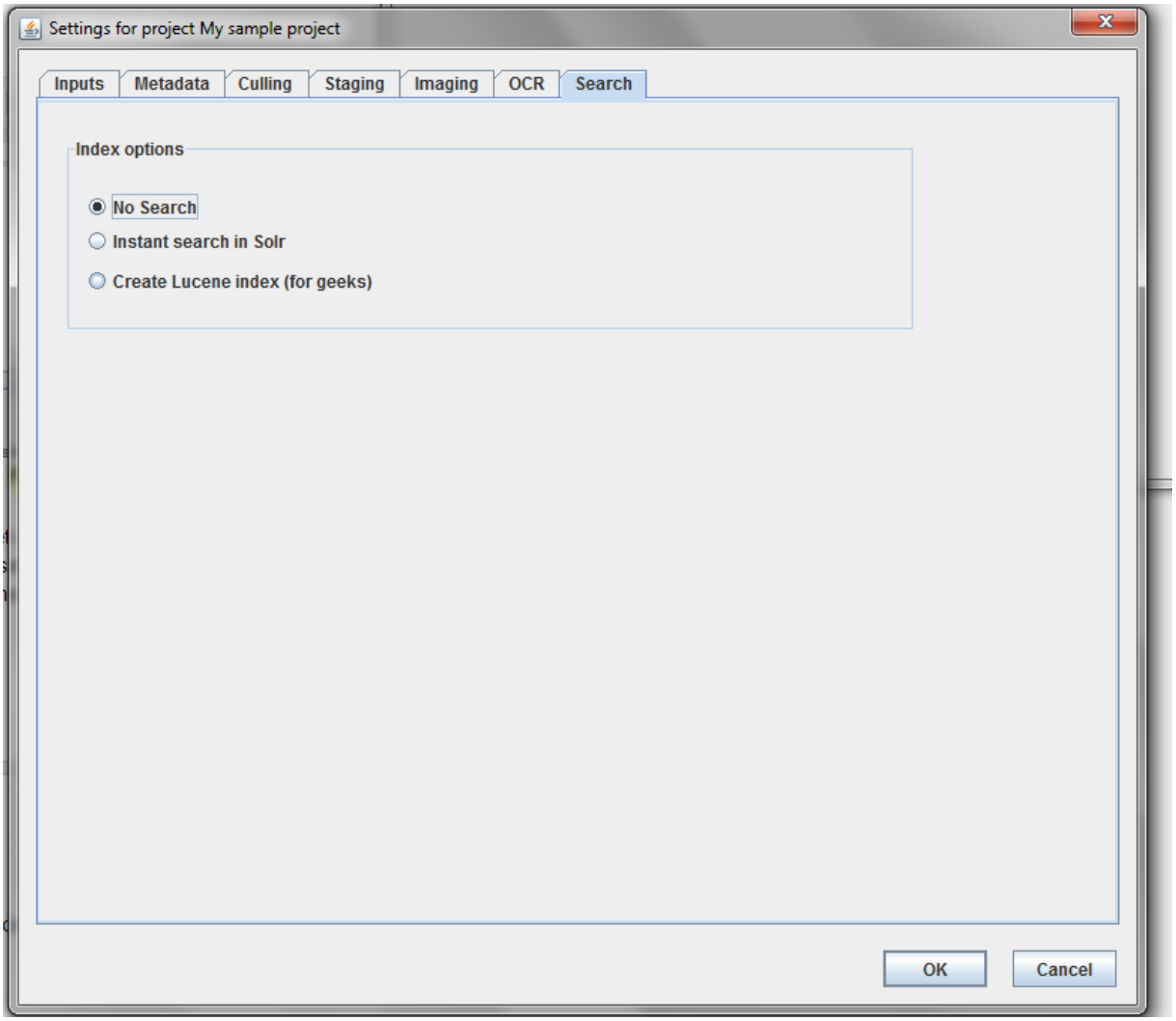
FIGURE 5.1

Since this is a sample project that was set up for the purpose of showing the user how things work, this figure is just for you to see the settings that are a part of the sample project. You can change the settings here, or just accept the existing ones by clicking OK. Later, when we are

setting up new projects, we will discuss how to make changes in this screen.

Before we can begin processing our sample project, we must make sure that some of our other basic settings are properly checked. Select the "**Search**" button as seen in the upper section of the Settings screen.

A Search screen, as seen below, will come up. Since we are simply running a test project to get the feel of things, make sure that "**No Search**" is selected in this window. Running your project with the search options turned on, will increase the processing time. For now we just want to learn how to use our SHMcloud™ Player. We will discuss the other options later on.



Note: If you already ran this project a few times, then each time the program is run it creates a new time-stamped folder to hold the results. In this case, you will first have to choose which "run" folder to open, or to create a new run. The timestamped "run" folder is created when you do staging.

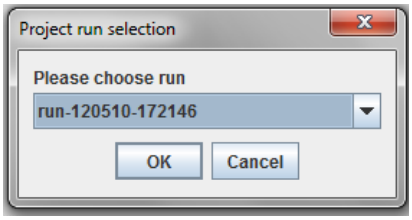


FIGURE 5.2

If this is not the first time you are running this test project, then a window similar to Figure 5.2 will open. Choose which run you would like, and click OK.

If this is the first time running the project, then nothing happens here and you may proceed to #6, Processing Your Test Job.

Note: You can remove any of the projects from a particular run by selecting them from the window seen above in Figure 5.1, and then clicking on the “Remove” button in the upper right side of the screeng.

6. Processing your test job

Now you are ready to Process this test job. Click on the Process Tab and select the “Stage” option as shown in Figure 6. If you are looking at your “Processing history” window, then you will see activity taking place when you select the “Stage” button. You may also see a bit of activity in the CMD window.

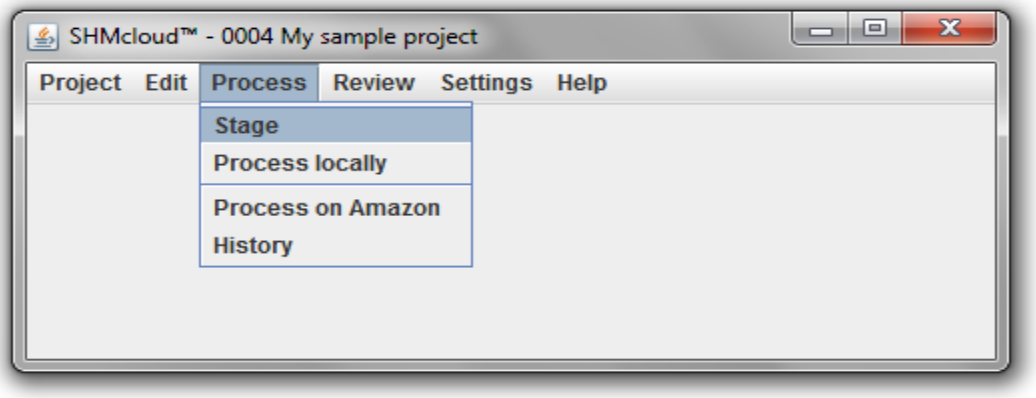


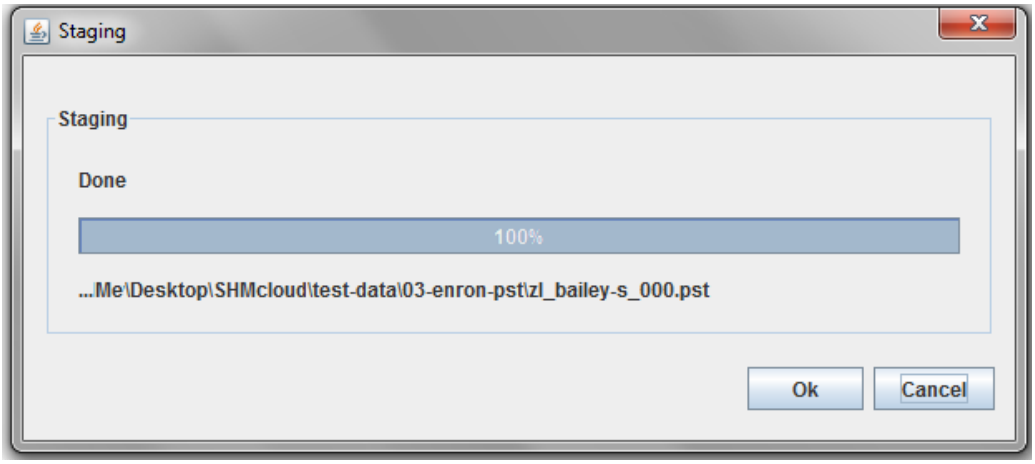
FIGURE 6

What is staging?

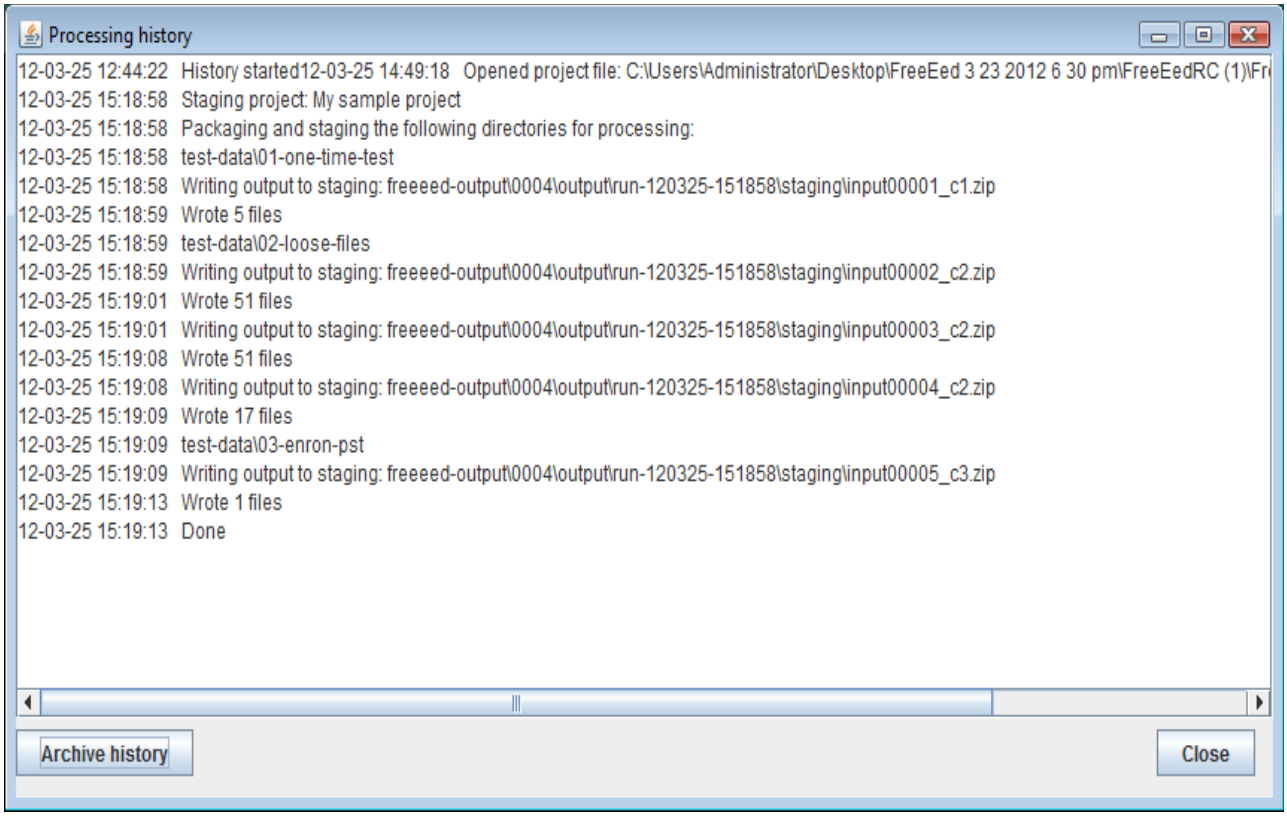
At this point the program combines all the input directories into zip files. It will use them for multiple purposes: to protect the original files, to break computation in stages, and in case of cloud computation - to upload these zip files to S3 (Amazon Simple Storage Solution), in order to process them in Hadoop.

Staging must be done before <u>any</u> project can be run.

As soon as you select the Staging option, a Screen will pop up showing you the progress of your staged project.



Once Staging has completed, simply push the Ok button to continue. You may also notice that the word “Done” appears in your History or Cmd windows.


Processing history
12-03-25 12:44:22 History started12-03-25 14:49:18 Opened project file: C:\Users\Administrator\Desktop\FreeEed 3 23 2012 6 30 pm\FreeEedRC (1)\Fr
12-03-25 15:18:58 Staging project: My sample project
12-03-25 15:18:58 Packaging and staging the following directories for processing:
12-03-25 15:18:58 test-data\01-one-time-test
12-03-25 15:18:58 Writing output to staging: freeeed-output\0004\output\run-120325-151858\staging\input00001_c1.zip
12-03-25 15:18:59 Wrote 5 files
12-03-25 15:18:59 test-data\02-loose-files
12-03-25 15:18:59 Writing output to staging: freeeed-output\0004\output\run-120325-151858\staging\input00002_c2.zip
12-03-25 15:19:01 Wrote 51 files
12-03-25 15:19:01 Writing output to staging: freeeed-output\0004\output\run-120325-151858\staging\input00003_c2.zip
12-03-25 15:19:08 Wrote 51 files
12-03-25 15:19:08 Writing output to staging: freeeed-output\0004\output\run-120325-151858\staging\input00004_c2.zip
12-03-25 15:19:09 Wrote 17 files
12-03-25 15:19:09 test-data\03-enron-pst
12-03-25 15:19:09 Writing output to staging: freeeed-output\0004\output\run-120325-151858\staging\input00005_c3.zip
12-03-25 15:19:13 Wrote 1 files
12-03-25 15:19:13 Done
Archive history
Close

7. Process Locally

Now that the data has been “Staged” you are ready to process the data. Pull down the Process Menu and select “Process locally” as shown in Figure 7.2.

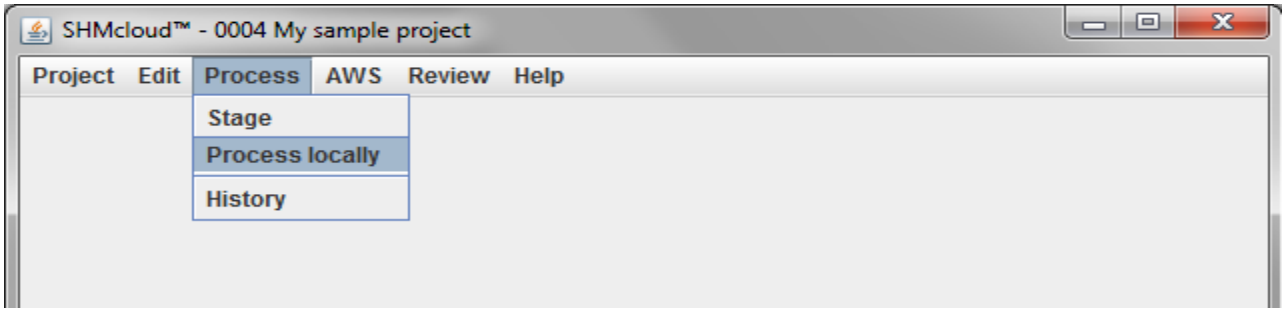


FIGURE 7.2

Note: If your data files are small enough, then you should have no problem processing your data locally. Processing your data locally takes full advantage of your free SHMcloud™ software without the Amazon interface or fees. Later in sections #11 and #12 we will learn how to process much larger files using the SHMcloud™ Player with an Amazon Web Service (AWS) account.

When the job is finished processing, your history window will look similar to Figure 7.3.

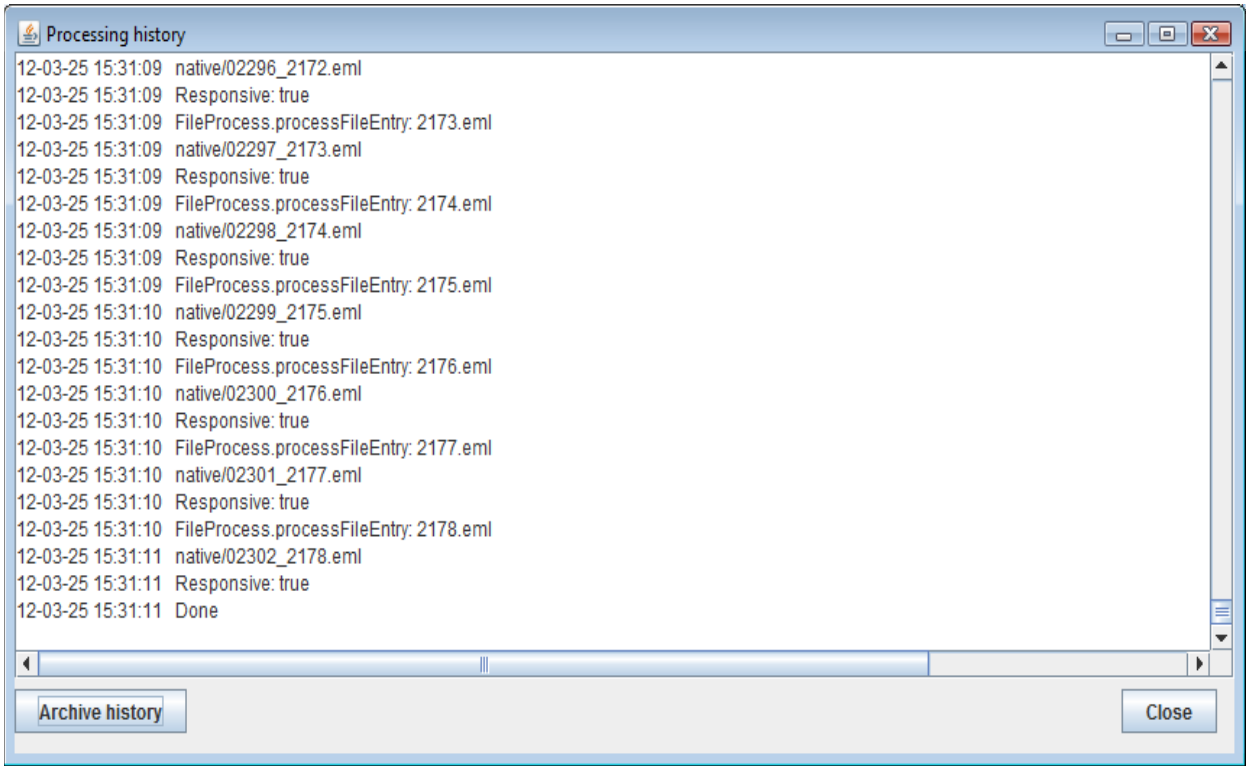


FIGURE 7.3

Your command window will also show some activity during the above process. This is normal, and is simply telling us that your Player is trying to process the data. When it is done, as in the example above, the word “Done” should appear.

Since no filtering has been added for the data, all documents were returned as True (vs False) when evaluated for being responsive. We will be discussing data Filtering in a later section.

8. Reviewing the results

Now we want to look at our output. To accomplish this task you will select “Open output folder” from the “Review” menu as shown in Figure 8.1.

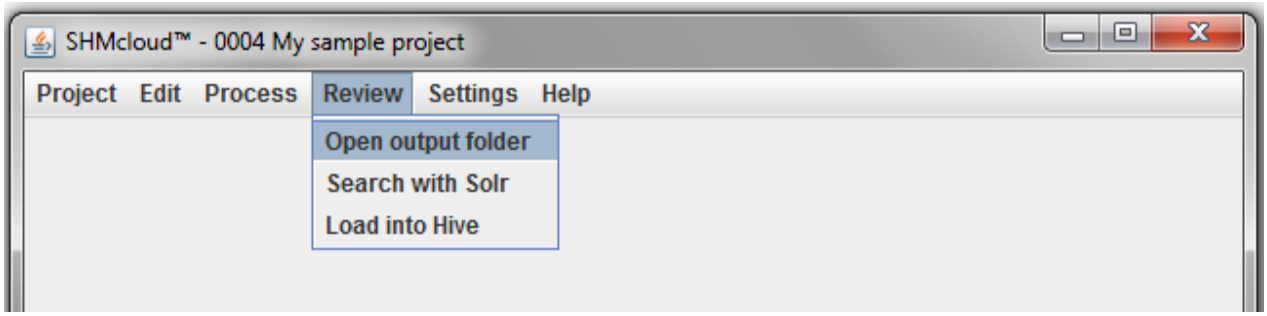


FIGURE 8.1

This action will bring up a window like Figure 8.2.

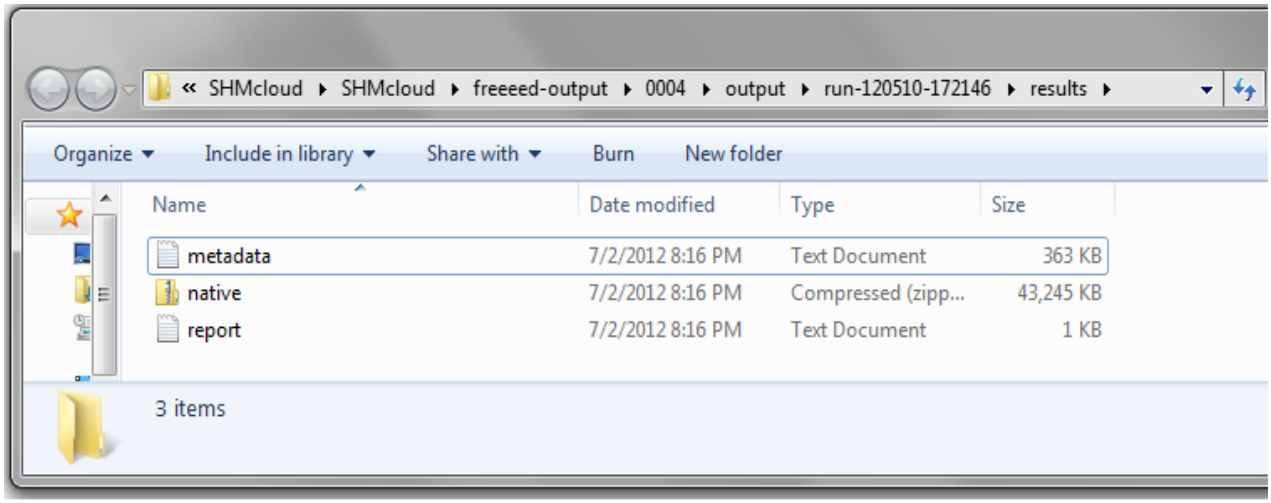


FIGURE 8.2

Note that you can manually drill down through the directories (starting from your SHMcloud™ directory) and get to the same data. The top folder is freeeed-output. The rest of the file path is displayed in Figure 8.2 above. Folder 0004/output folder/run-XXX folder/ is the folder that contains the results from this particular sample test project.

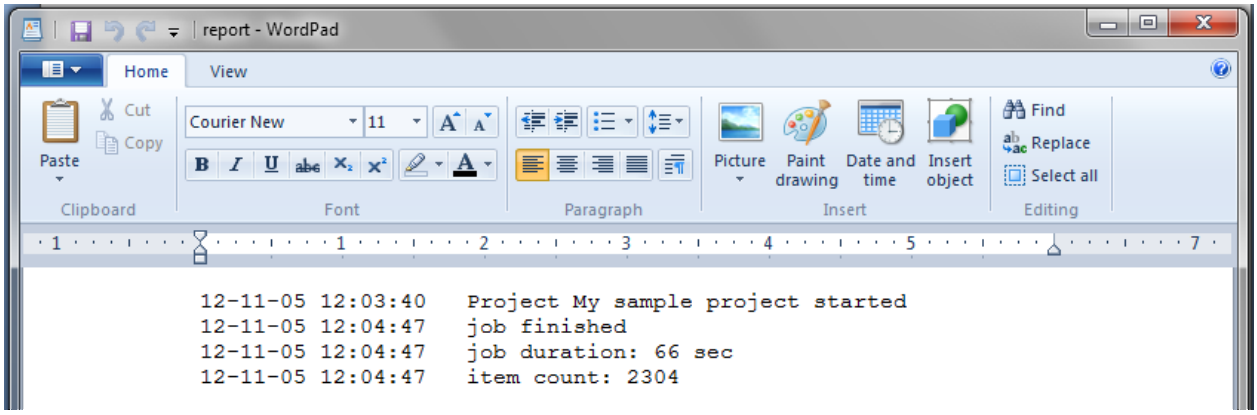
Each time you process a job in SHMcloud™ a new folder will be generated for storing your

output, as well as your original data. You will be able to access those output folders by opening the corresponding project from within your SHMcloud™ Player, or by simply drilling down directly from you SHMcloud folder, and through to your freeeed-output folder.

Also note that if you open the zipped native folder you will find a variety of file types that were processed by SHMcloud™ including mail, PDF, Excel, PowerPoint etc. We will discuss all of those folders shortly.

. Report

Clicking on the Report folder will render results similar to the following image.



The Report file only gets produced when you are running your project Locally. This file will not be created if you run a project on Amazon.

Additionally, if your project terminated prematurely, the Report file will also fail to be produced. The Report files will only be created if your project was successfully run.

This particular file is telling us that the data processed in only 66 seconds, and the entire output consists of a total of 2304 files, records, images, etc.

9. Metadata

The Metadata file is akin to a very detailed index. It consists of the names of every file that is run through your project, regardless of whether or not your SHMcloud™ Player is able to process it. The Metadata includes the names of corresponding Custodians for each file, as well as any other detailed information that is relevant to that file. When you begin working with Searching, the Metadata file can be a very useful tool for helping you to pinpoint your Searches.

9.1. Now it is time to take a look at the metadata file that SHMcloud™ created. To view the data you can use Excel™ or Open Office Calc. I have chosen Open Office Calc to display the data. Right click on the “metada” file , slide down the menu that appears to select “Open with” then slide to the right and select the program to view the metadata file with, in my case I am using Open Office Calc, as shown in Figure 9.1.

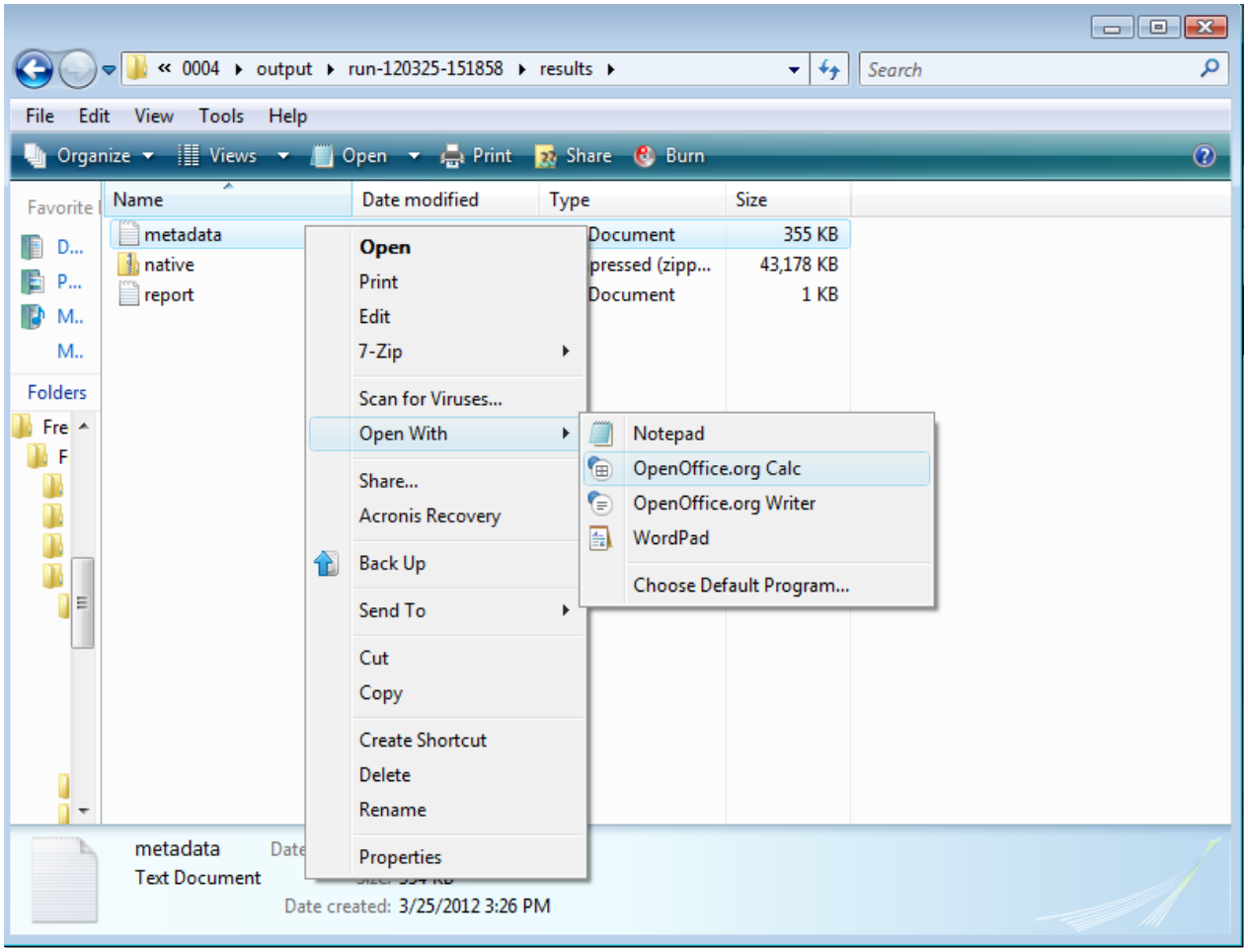


FIGURE 9.1

9.2. When you are opening the data you will need to select “Other” and the delimiter needs to be “pipe” which is the key above the “Enter” key on the keyboard which is entered while holding down the shift key, and will look like Figure 9.2. (The box next to “Other” should contain the aforementioned pipe “|”.)

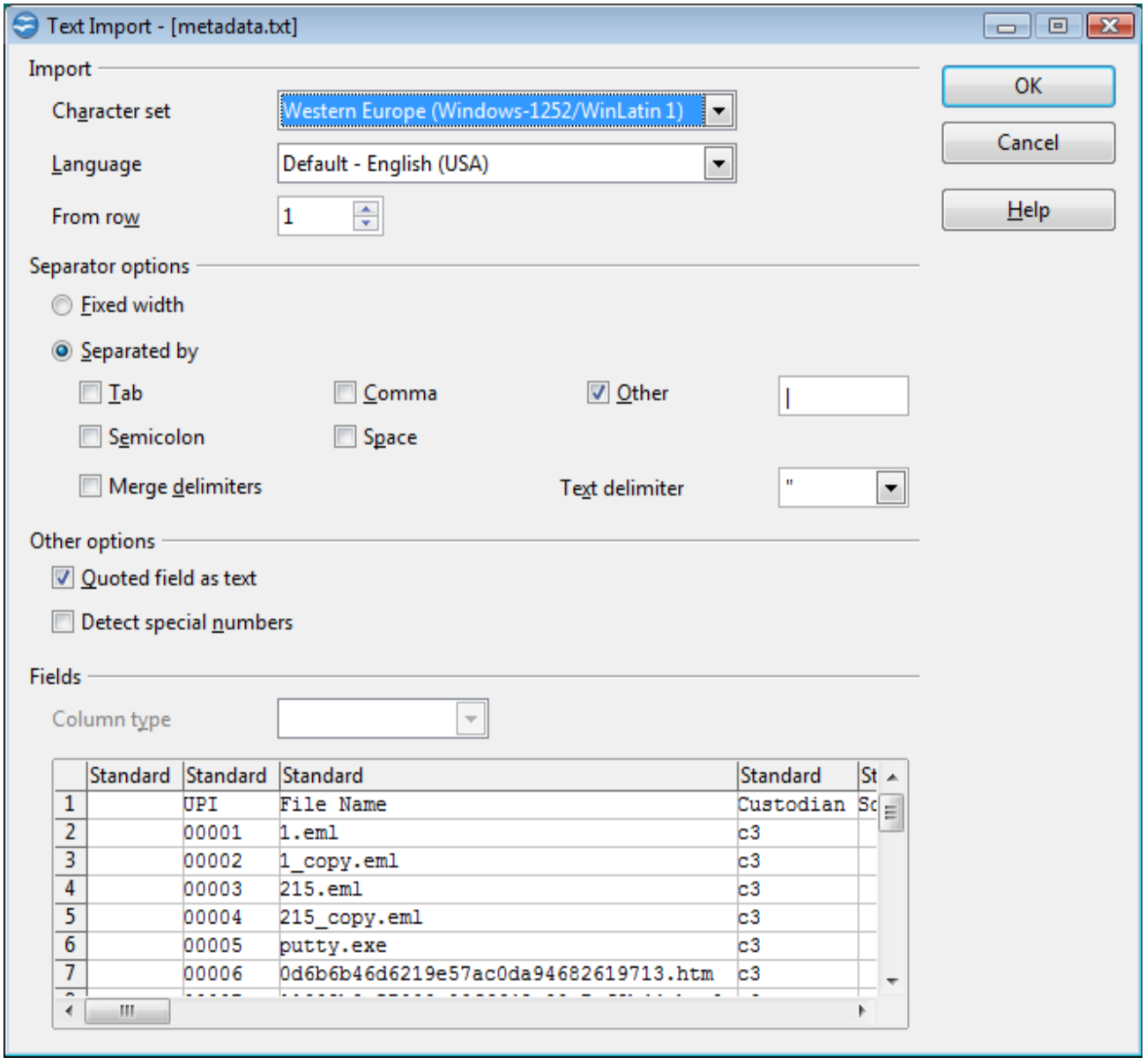


FIGURE 9.2

9.3 Standard metadata fields.

SHMcloud extracts the metadata fields and names them according to the industry standard. The names and their aliases are set in the file

config/standard-metadata-names.properties

By changing these, you can make SHMcloud call the fields differently, or extract different fields under different names. Here is the default content of this file

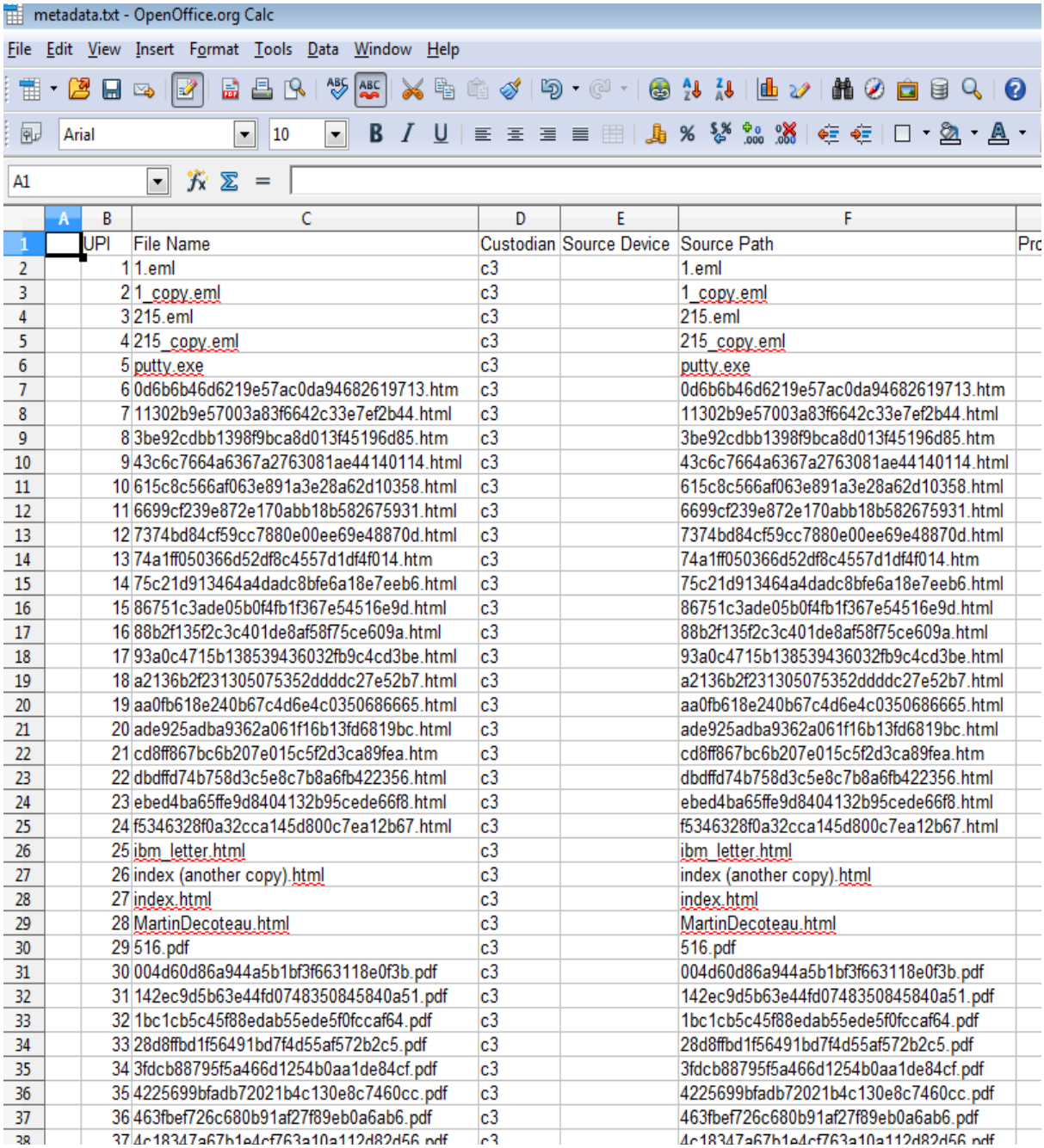
```
# Based on Judge Shira Sheindlin decision:
#
http://scholar.google.com/scholar_case?case=14703320529971186199&hl=en&as_sdt=2&as_vis=1&oi=scholarr

# First mentioned is the standard name, it is better not be changed, unless you know what you are doing.
# Following names separated by commas are variants, or aliases, found in native metadata, to be mapped to this name.


01=UPI
02=File Name
03=Custodian
04=Source Device
05=Source Path, document_original_path
06=Production Path
07=Modified Date
08=Modified Time
09=Time Offset Value
10=processing_exception
11=master_duplicate
12=text

#Additional fields e-mail messages.

21=To, Message-To
22=From, Author, Message-From
23=CC, Message-Cc
24=BCC, Message-Bcc
25=Date Sent
```

26=Time Sent
27=Subject, subject
28=Date Received, date
29=Time Received
#Attachments: The Bates number ranges of e-mail attachments.
#The parties may alternatively choose to use: Bates_Begin, Bates_End, Attach_Begin and Attach_End.

#Helpfule artifacts
31=native_link
32=text_link
33=exception_link

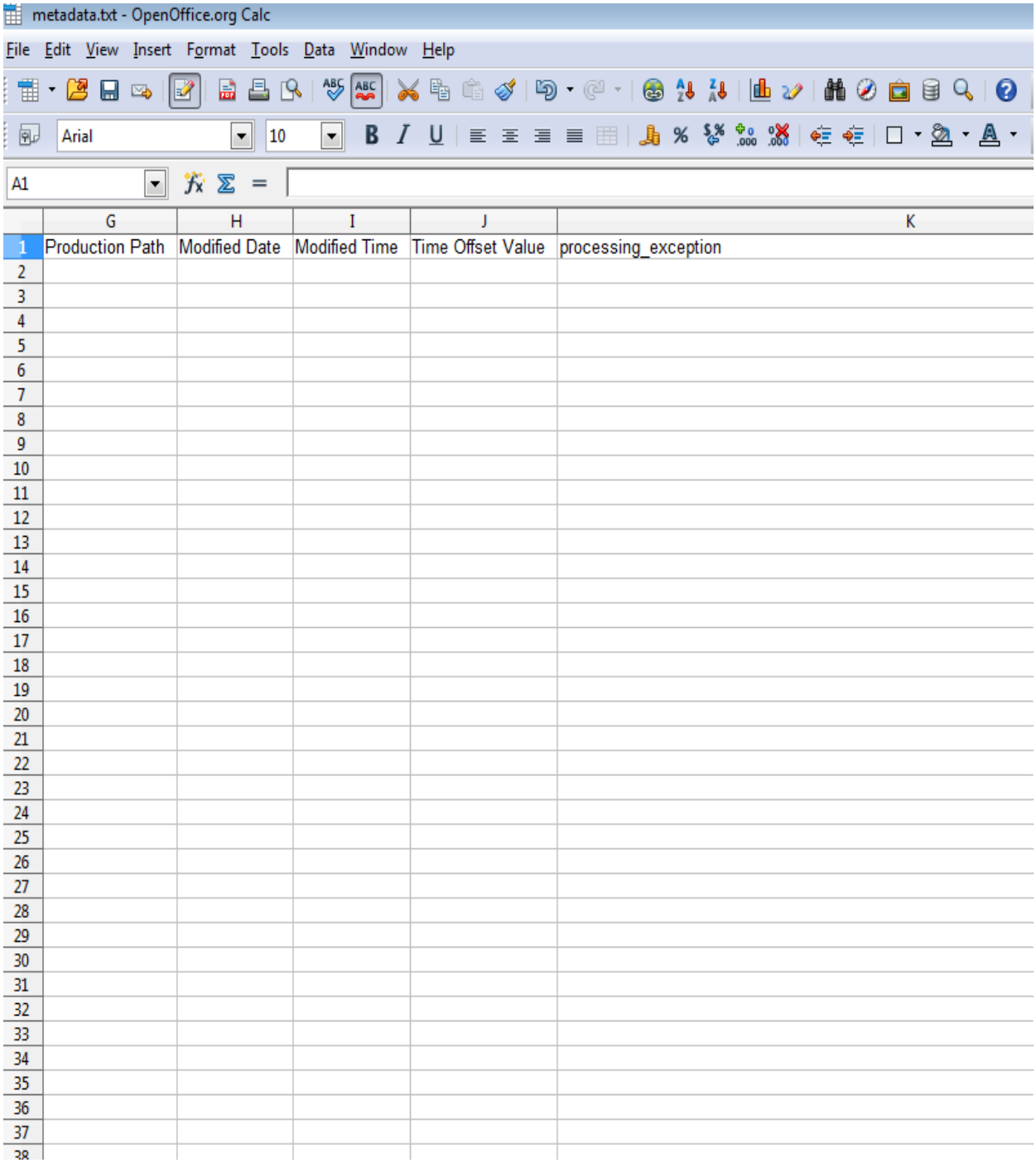
Now let us take a look at the metadata fields that SHMCloud™ created. Starting at the upper left of the table and moving from left to right, we can see the various metadata fields created by the processing. As shown below in Figures 9.3a through 9.3e, the output produces a report with many different fields. In Section 10 we will be discussing how to create and save your own projects. Part of creating your project will be to assign a “custodian” to the different files that you will be processing. Please note the custodian field below and how it relates to each line of output.

As you can see, in essence the metadata file is a list of all the records that our project has processed, including relevant information pertaining to those records to aid you in your detailed searches. We will be discussing different search options shortly.



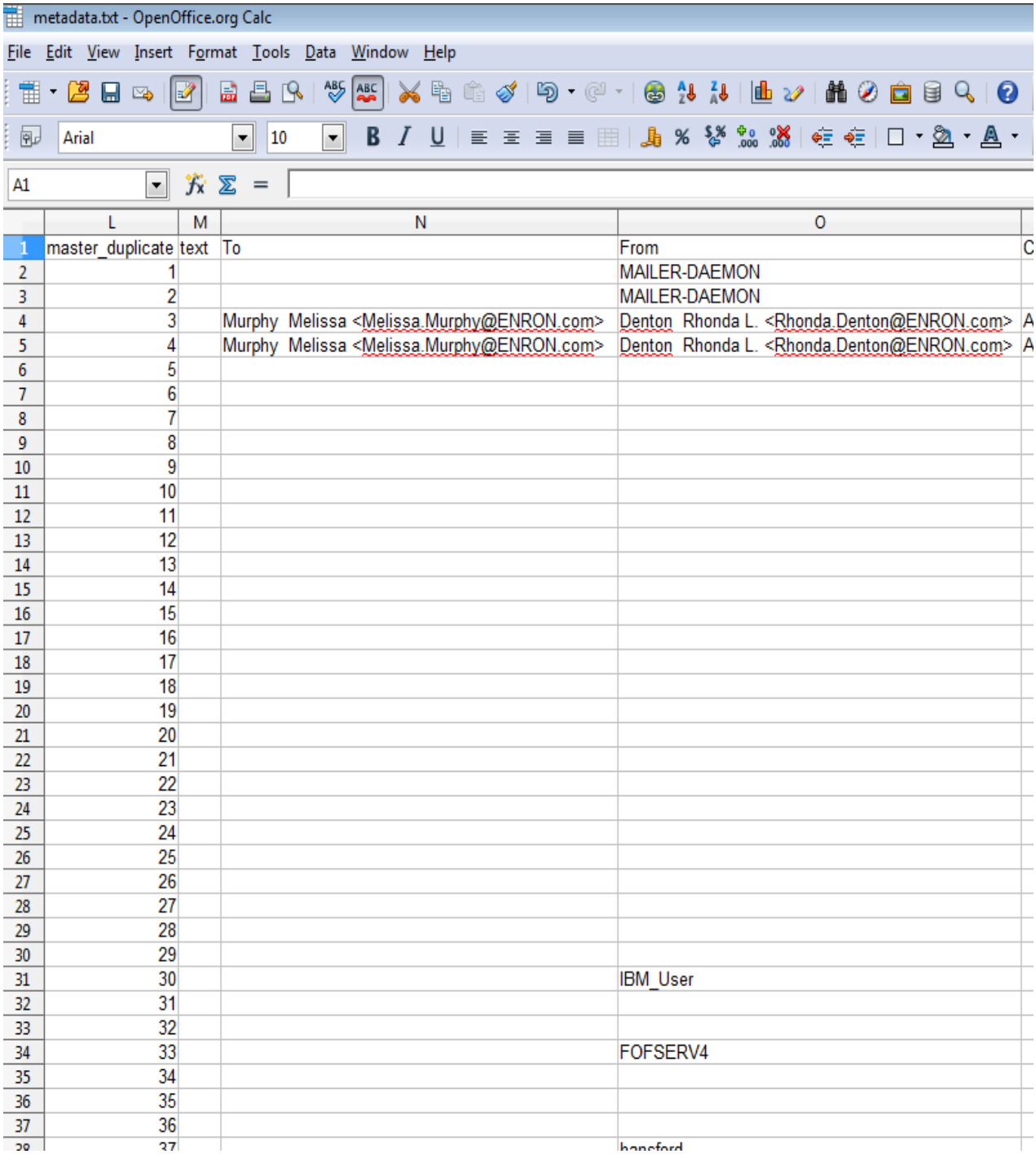
	A	B	C	D	E	F	
1		UPI	File Name	Custodian	Source Device	Source Path	Pro
2		1	1.eml	c3		1.eml	
3		2	1_copy.eml	c3		1_copy.eml	
4		3	215.eml	c3		215.eml	
5		4	215_copy.eml	c3		215_copy.eml	
6		5	putty.exe	c3		putty.exe	
7		6	0d6b6b46d6219e57ac0da94682619713.htm	c3		0d6b6b46d6219e57ac0da94682619713.htm	
8		7	11302b9e57003a83f6642c33e7ef2b44.html	c3		11302b9e57003a83f6642c33e7ef2b44.html	
9		8	3be92cdbb1398f9bca8d013f45196d85.htm	c3		3be92cdbb1398f9bca8d013f45196d85.htm	
10		9	43c6c7664a6367a2763081ae44140114.html	c3		43c6c7664a6367a2763081ae44140114.html	
11		10	615c8c566af063e891a3e28a62d10358.html	c3		615c8c566af063e891a3e28a62d10358.html	
12		11	6699cf239e872e170abb18b582675931.html	c3		6699cf239e872e170abb18b582675931.html	
13		12	7374bd84cf59cc7880e00ee69e48870d.html	c3		7374bd84cf59cc7880e00ee69e48870d.html	
14		13	74a1ff050366d52df8c4557d1df4f014.htm	c3		74a1ff050366d52df8c4557d1df4f014.htm	
15		14	75c21d913464a4dadc8bfe6a18e7eeb6.html	c3		75c21d913464a4dadc8bfe6a18e7eeb6.html	
16		15	86751c3ade05b0f4fb1f367e54516e9d.html	c3		86751c3ade05b0f4fb1f367e54516e9d.html	
17		16	88b2f135f2c3c401de8af58f75ce609a.html	c3		88b2f135f2c3c401de8af58f75ce609a.html	
18		17	93a0c4715b138539436032fb9c4cd3be.html	c3		93a0c4715b138539436032fb9c4cd3be.html	
19		18	a2136b2f231305075352ddddc27e52b7.html	c3		a2136b2f231305075352ddddc27e52b7.html	
20		19	aa0fb618e240b67c4d6e4c0350686665.html	c3		aa0fb618e240b67c4d6e4c0350686665.html	
21		20	ade925adba9362a061f16b13fd6819bc.html	c3		ade925adba9362a061f16b13fd6819bc.html	
22		21	cd8ff867bc6b207e015c5f2d3ca89fea.htm	c3		cd8ff867bc6b207e015c5f2d3ca89fea.htm	
23		22	dbdffd74b758d3c5e8c7b8a6fb422356.html	c3		dbdffd74b758d3c5e8c7b8a6fb422356.html	
24		23	ebed4ba65ffe9d8404132b95cede66f8.html	c3		ebed4ba65ffe9d8404132b95cede66f8.html	
25		24	f5346328f0a32cca145d800c7ea12b67.html	c3		f5346328f0a32cca145d800c7ea12b67.html	
26		25	ibm_letter.html	c3		ibm_letter.html	
27		26	index (another copy).html	c3		index (another copy).html	
28		27	index.html	c3		index.html	
29		28	MartinDecoteau.html	c3		MartinDecoteau.html	
30		29	516.pdf	c3		516.pdf	
31		30	004d60d86a944a5b1bf3f663118e0f3b.pdf	c3		004d60d86a944a5b1bf3f663118e0f3b.pdf	
32		31	142ec9d5b63e44fd0748350845840a51.pdf	c3		142ec9d5b63e44fd0748350845840a51.pdf	
33		32	1bc1cb5c45f88edab55ede5f0fccaf64.pdf	c3		1bc1cb5c45f88edab55ede5f0fccaf64.pdf	
34		33	28d8ffbd1f56491bd7f4d55af572b2c5.pdf	c3		28d8ffbd1f56491bd7f4d55af572b2c5.pdf	
35		34	3fdcb88795f5a466d1254b0aa1de84cf.pdf	c3		3fdcb88795f5a466d1254b0aa1de84cf.pdf	
36		35	4225699bfadb72021b4c130e8c7460cc.pdf	c3		4225699bfadb72021b4c130e8c7460cc.pdf	
37		36	463fbef726c680b91af27f89eb0a6ab6.pdf	c3		463fbef726c680b91af27f89eb0a6ab6.pdf	
38		37	4c18347a67b1e4cf763a10a112d82d56.pdf	c3		4c18347a67b1e4cf763a10a112d82d56.pdf	

FIGURE 9.3a



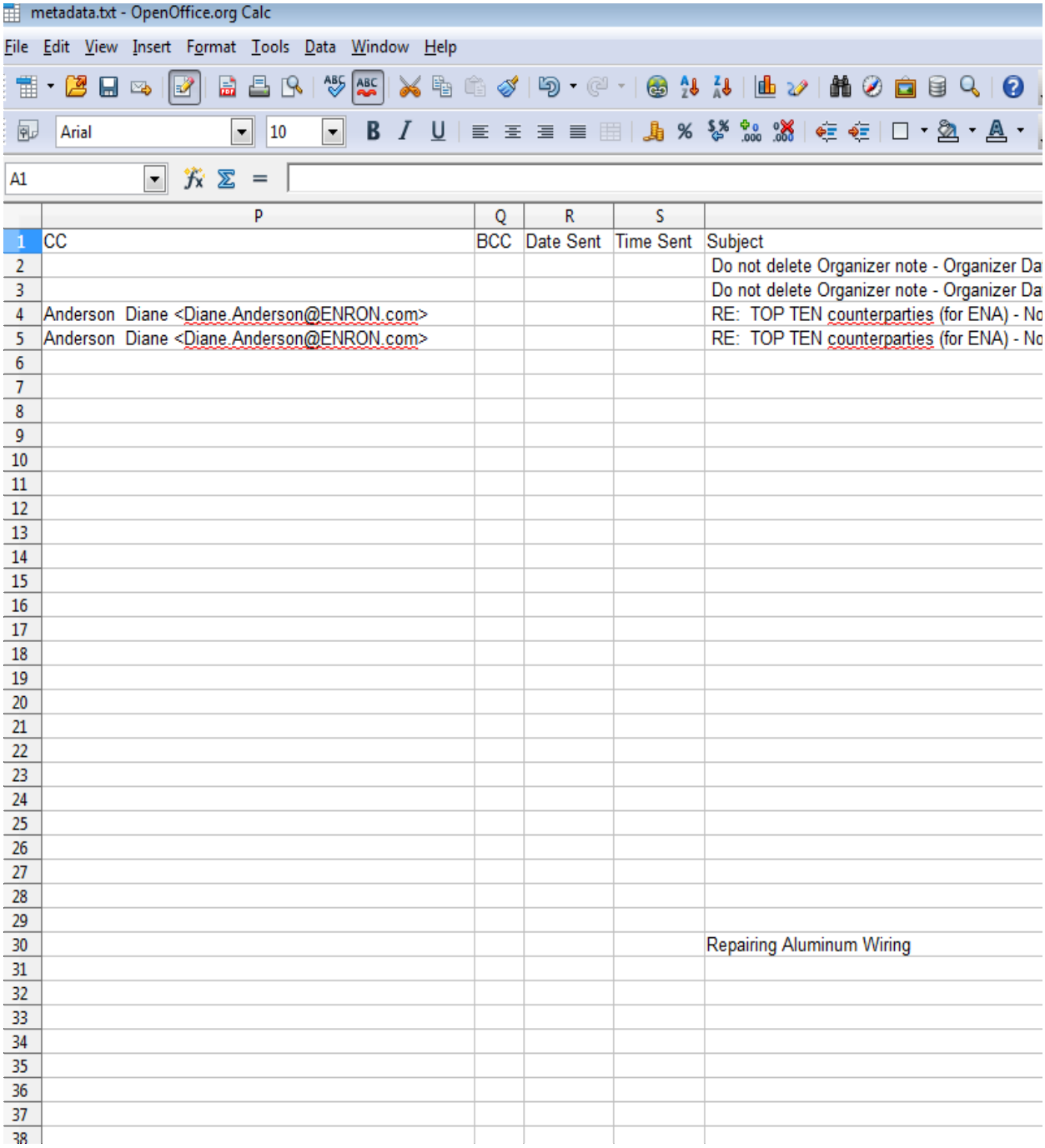
	G	H	I	J	K
1	Production Path	Modified Date	Modified Time	Time Offset Value	processing_exception
2					
3					
4					
5					
6					
7					
8					
9					
10					
11					
12					
13					
14					
15					
16					
17					
18					
19					
20					
21					
22					
23					
24					
25					
26					
27					
28					
29					
30					
31					
32					
33					
34					
35					
36					
37					

FIGURE 9.3b



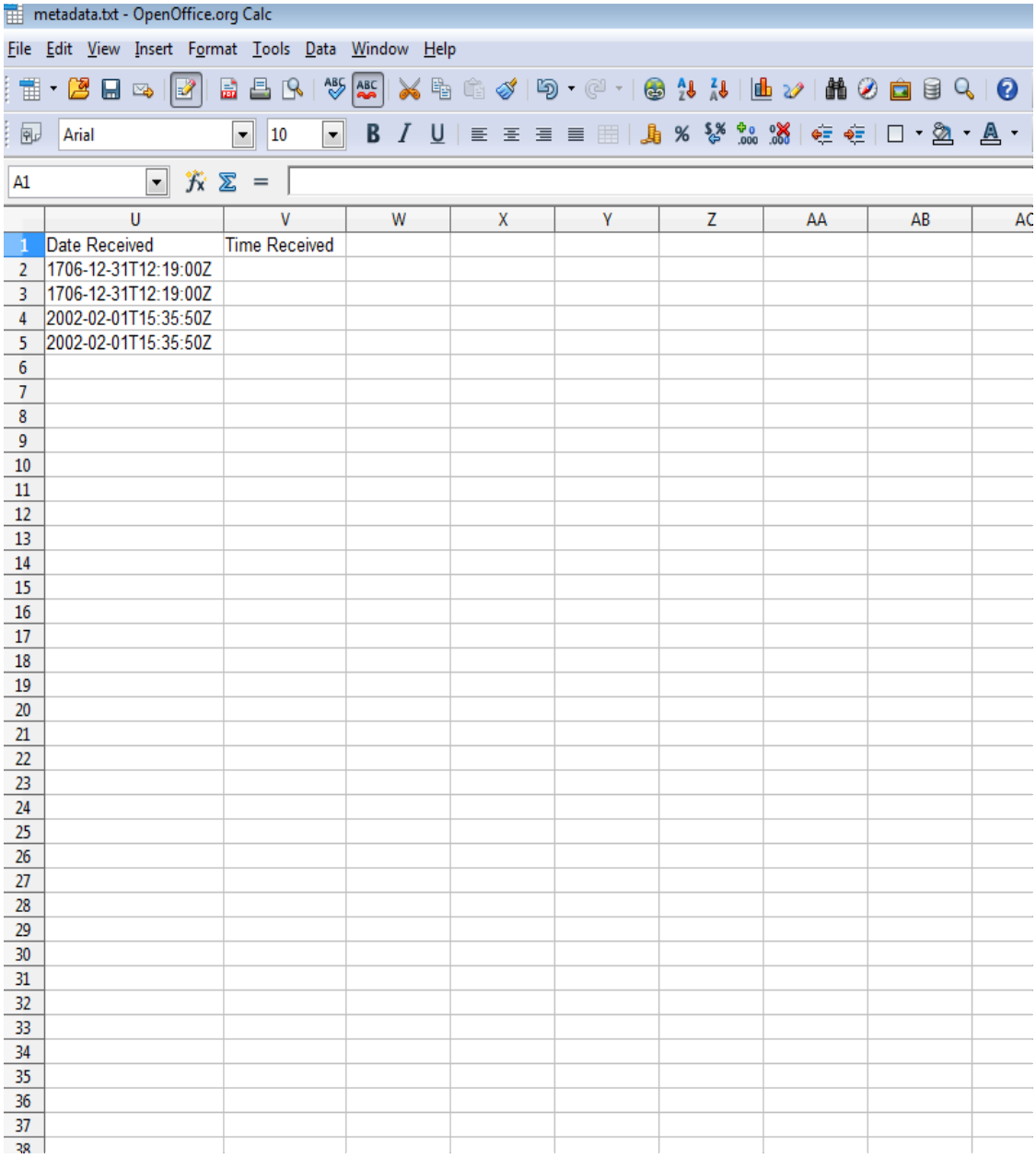
	L	M	N	O	
1	master_duplicate	text	To	From	C
2	1			MAILER-DAEMON	
3	2			MAILER-DAEMON	
4	3		Murphy Melissa <Melissa.Murphy@ENRON.com>	Denton Rhonda L. <Rhonda.Denton@ENRON.com>	A
5	4		Murphy Melissa <Melissa.Murphy@ENRON.com>	Denton Rhonda L. <Rhonda.Denton@ENRON.com>	A
6	5				
7	6				
8	7				
9	8				
10	9				
11	10				
12	11				
13	12				
14	13				
15	14				
16	15				
17	16				
18	17				
19	18				
20	19				
21	20				
22	21				
23	22				
24	23				
25	24				
26	25				
27	26				
28	27				
29	28				
30	29				
31	30			IBM_User	
32	31				
33	32				
34	33			FOFSERV4	
35	34				
36	35				
37	36				
38	37			hansford	

FIGURE 9.3c



	P	Q	R	S	
1	CC	BCC	Date Sent	Time Sent	Subject
2					Do not delete Organizer note - Organizer Da
3					Do not delete Organizer note - Organizer Da
4	Anderson Diane <Diane.Anderson@ENRON.com>				RE: TOP TEN counterparties (for ENA) - No
5	Anderson Diane <Diane.Anderson@ENRON.com>				RE: TOP TEN counterparties (for ENA) - No
6					
7					
8					
9					
10					
11					
12					
13					
14					
15					
16					
17					
18					
19					
20					
21					
22					
23					
24					
25					
26					
27					
28					
29					
30					Repairing Aluminum Wiring
31					
32					
33					
34					
35					
36					
37					
38					

FIGURE 9.3d

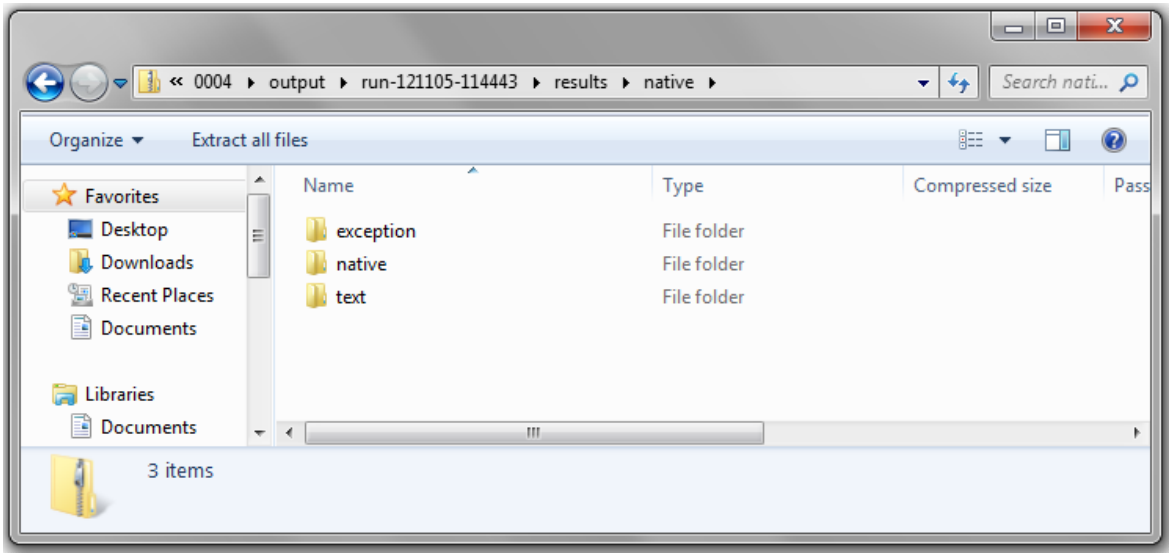


	U	V	W	X	Y	Z	AA	AB	AC
1	Date Received	Time Received							
2	1706-12-31T12:19:00Z								
3	1706-12-31T12:19:00Z								
4	2002-02-01T15:35:50Z								
5	2002-02-01T15:35:50Z								
6									
7									
8									
9									
10									
11									
12									
13									
14									
15									
16									
17									
18									
19									
20									
21									
22									
23									
24									
25									
26									
27									
28									
29									
30									
31									
32									
33									
34									
35									
36									
37									
38									

FIGURE 9.3e

. Native Zip Folder

Clicking into your Native Zip Folder in the output section of your SHMcloud Player will reveal several different folders. These folders consists of our actual output, as well as the original input data.



In our current example we can see a folder called “exception”, another called “native” and another called “text”.

If we would have run this project with PDF imaging enabled, then we would have seen another folder called “pdf”. However, running a project with pdf imaging enabled, increases the run time considerably, and so we chose not to do that with our current test data since we are simply trying to learn how to use our Player.

We will now briefly discuss these three outputted folders. However, since they are currently located within a Zip folder, we advise you to extract the folders first. While it is possible to open files from withing a zipped folder, the results are not always complete or accurate.

You can simply copy these three folders someplace else on your hard drive or external storage device, perhaps in a folder with a distinctive name for your own personal use.

Exception Folder

The Exception Folder does not always get created during a project run. It will only get created in the event that your inputted data contains something that cannot be processed by the

SHMcloud™ player.

When you open the Exception folder you will see at least one document that could not be processed. If all of your records were able to be processed without exception, then the Exception Folder would not have been created.

You can easily access the data from within the exception folder. It is even possible that the Player will have processed your Exception files, however, this folder is bringing to your attention that there is something unusual about those particular files.

Native Folder

The Native folder contains all of the data that you put into this project. It is a folder that combines every file and every folder from every Custodian that was processed by this project, including any Exception files.

If you want to refer back to any of the files processed by your project in their original form, this is the folder to look at.

Text Folder

The Text folder is created automatically with every properly processed project. Each file that runs through the player is converted into a txt file. The text file is then placed into the Text folder.

⇛ If a character is unrecognized by the player, it might be replaced by a ? in the converted txt file.

We will discuss additional features pertaining to text conversions in a little while.

***Congratulations**! At this point we have done a complete run through of the provided test data set.*

10. Creating & saving your own project

Now we will cover the creation of a new project and get down to the business of processing your files.

First you must have data that you will be processing either on your computer or available to you from an external source. In most cases you will want to process an entire folder, but a single file can also be processed. Once you know where your data is being stored, then you will be able to run your own project. Similar to the test project that we processed earlier in this manual, we recommend that you create a Test Folder on your computer with a small amount of data that you can use to test out our software locally. We recommend that you try this before running your project using our Cloud processor through Amazon. As discussed earlier, local processing is free, but Cloud processing is not.

We start at the Project menu and select New from the Project menu as seen in Figure 10.1.

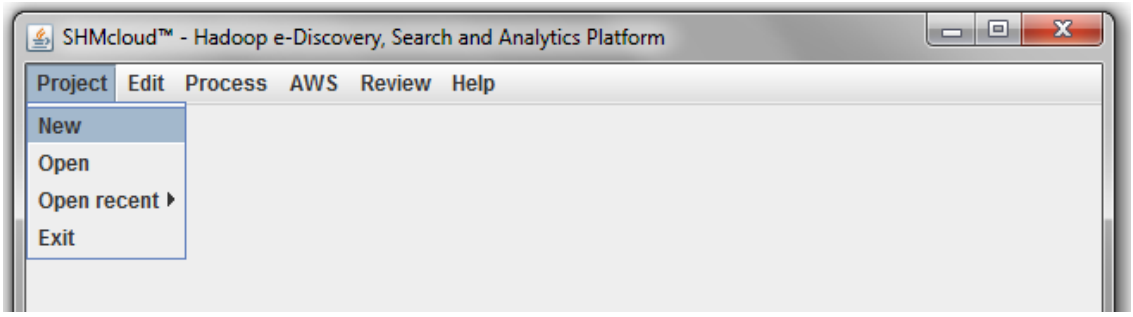


FIGURE 10.1

After starting a new project, provide a unique “Description” for easy identification in the future (Figure 10.2).

Please note: The name that you give for the Description of any project will appear at the top of every screen in the Title Bar after the project is saved and reopened - for every run of that same project in the future. So choose wisely!

Then click on “Add local folder” to select local documents, or select “Add network location” for files located on an intranet or on the internet.

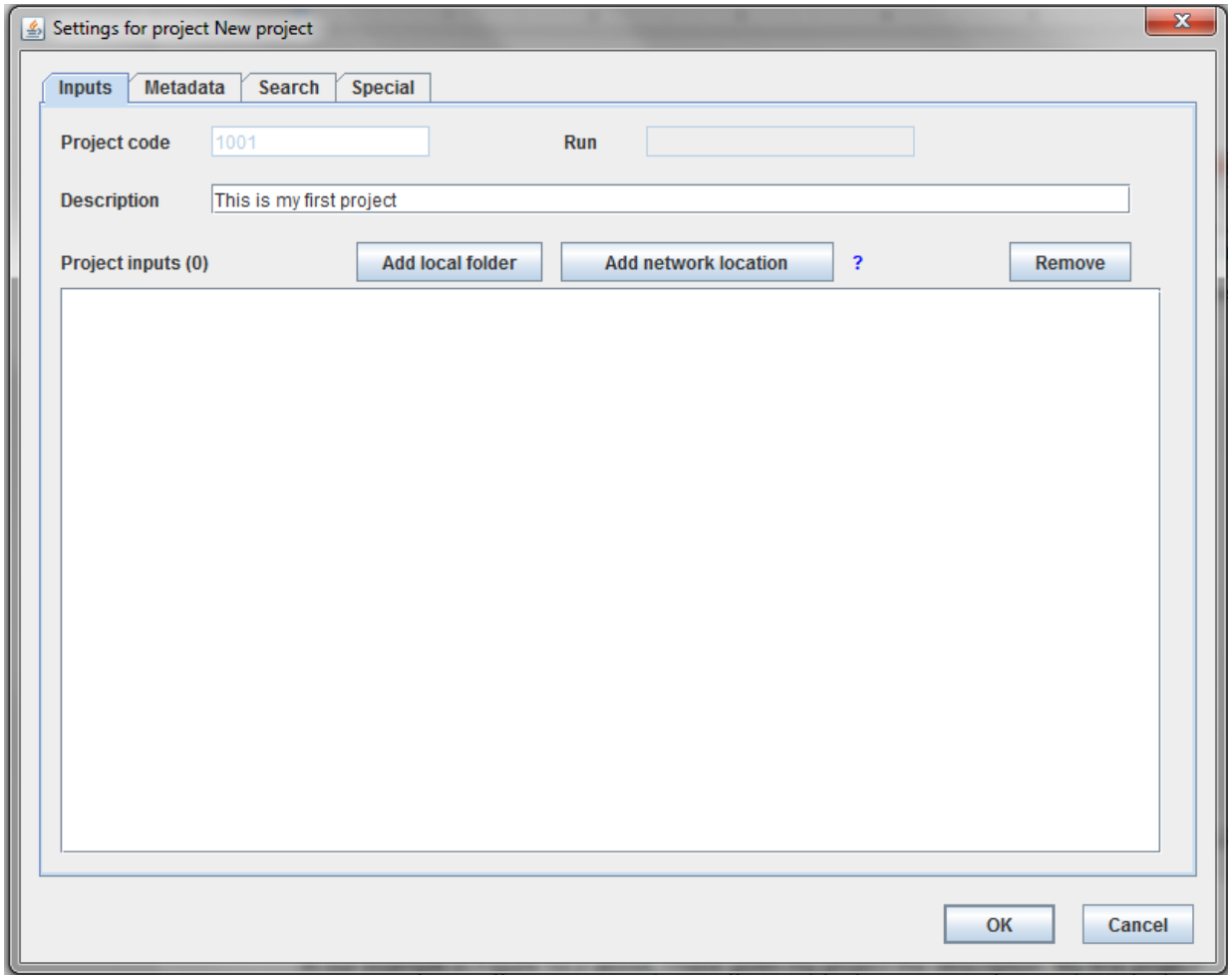


FIGURE 10.2

In our example in Figure 10.2 above, I have given my project the description “This is my first project”. You can give your project any description that you desire.

We recommend that you run a sample test of your own at this point, just to get the hang of it. If you do not have any test data to play with but would like to continue testing our product, SHMcloud™ provides our users with sample test data to use for running test.

Clicking on “Add local folder” will bring up a typical navigation window for selecting files or folders and will resemble Figure 10.3 below. You can select your data from within that directory, or use the navigation bar in the “Open” screen to access files from anyplace on your computer.

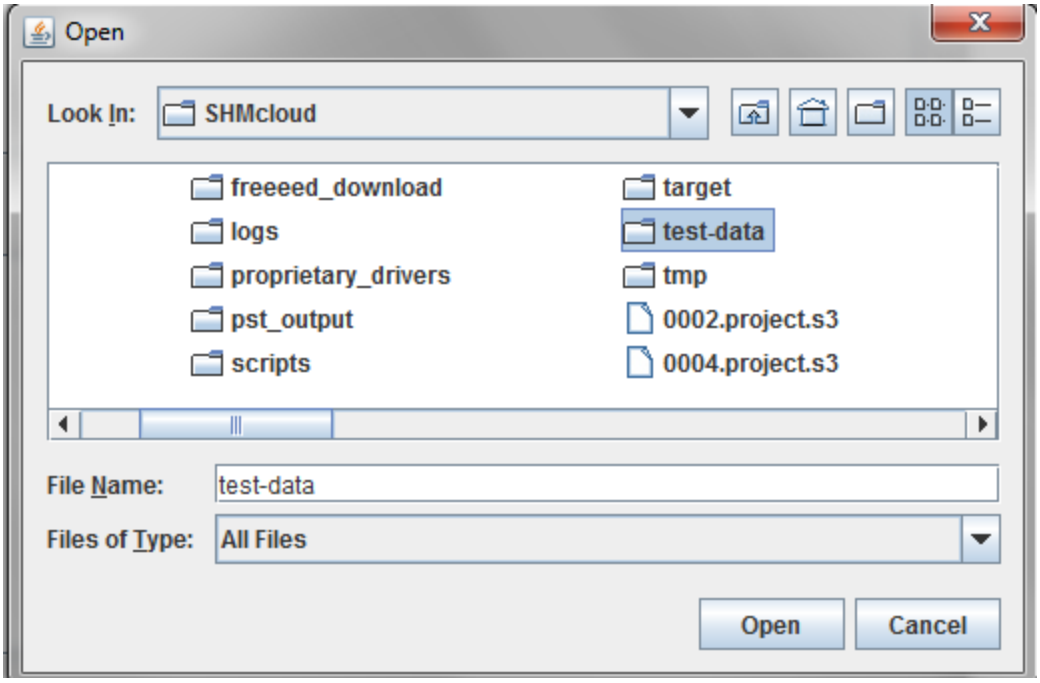


FIGURE 10.3

As mentioned, at this point you can choose to use the SHMcloud™ test data, and may do so by clicking on the “test-data” folder that is located within the SHMcloud™ directory, as seen in Figure 10.3 above.

I happen to have my own test folder that I would like to use. Here are the steps that I have taken in order to access my personal data:

In example 10.4 below, I selected “Add local folder”. Then I browsed through to my desktop where I happen to have a folder filled with Test data. The files that I would like to run through my SHMcloud™ software are located in my “Test data” folder.

I clicked on “Test data” and then “Open”.

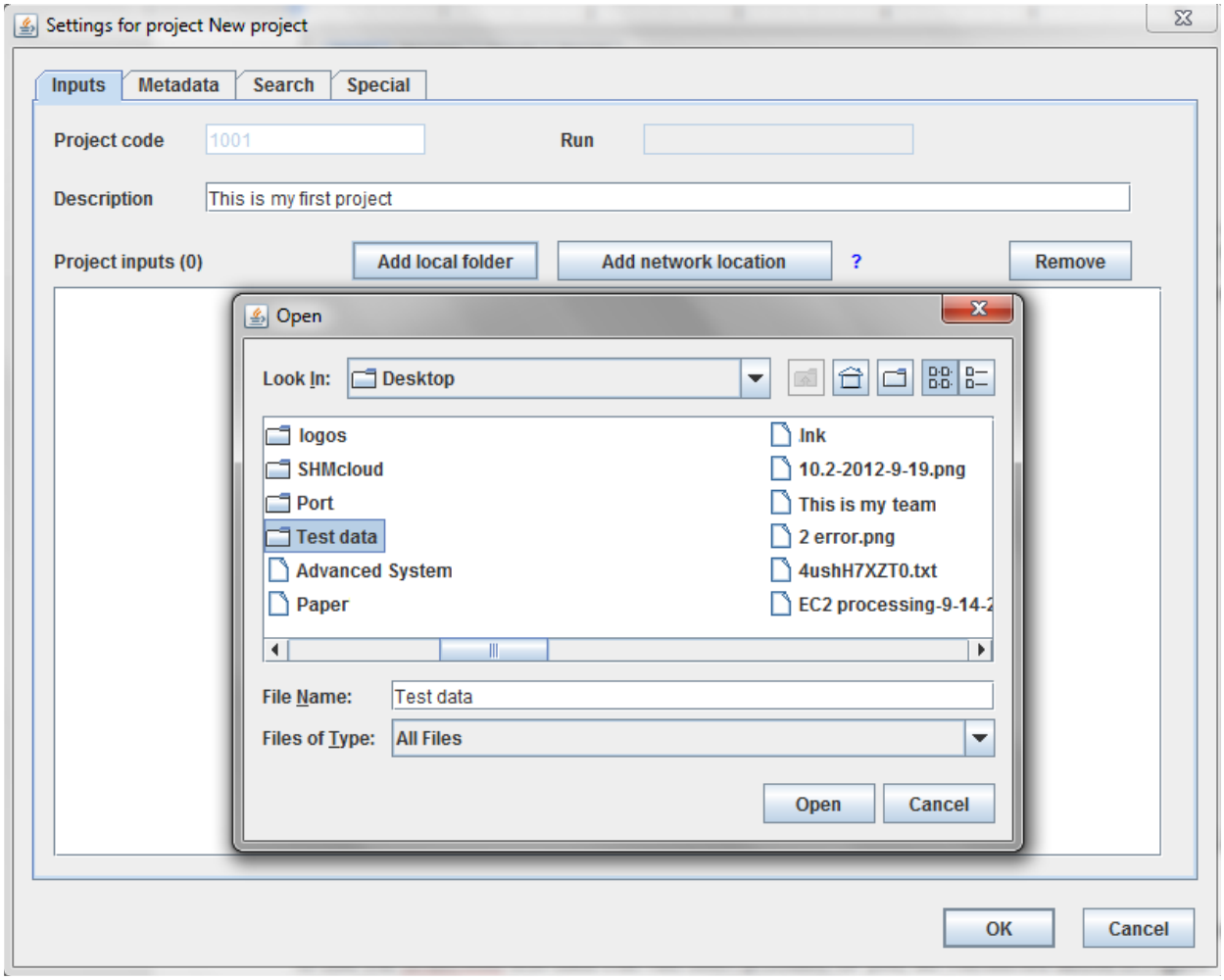


FIGURE 10.4

Do you have a directory filled with all kinds of files that you would like to run through the SHMcloud™ software? At this point you may open your own folder filled with files, or just choose to use the SHMcloud test-data that has been provided for you, as mentioned above in Figure 10.3. We recommend that you use a small folder at this point so that you can run a quick test.

Regardless of whether you are using the test data that SHMcloud™ provides, or if you are using your own test data, clicking “Open” will cause a dialog box to pop up asking the user to assign a “custodian’s name”, as seen in Figure 10.5 below.

The **custodian** defines whose files are being processed by the project. Later when you are processing massive amounts of data, this feature will be quite useful, as you can have many folders and different custodians being processed in the same project.

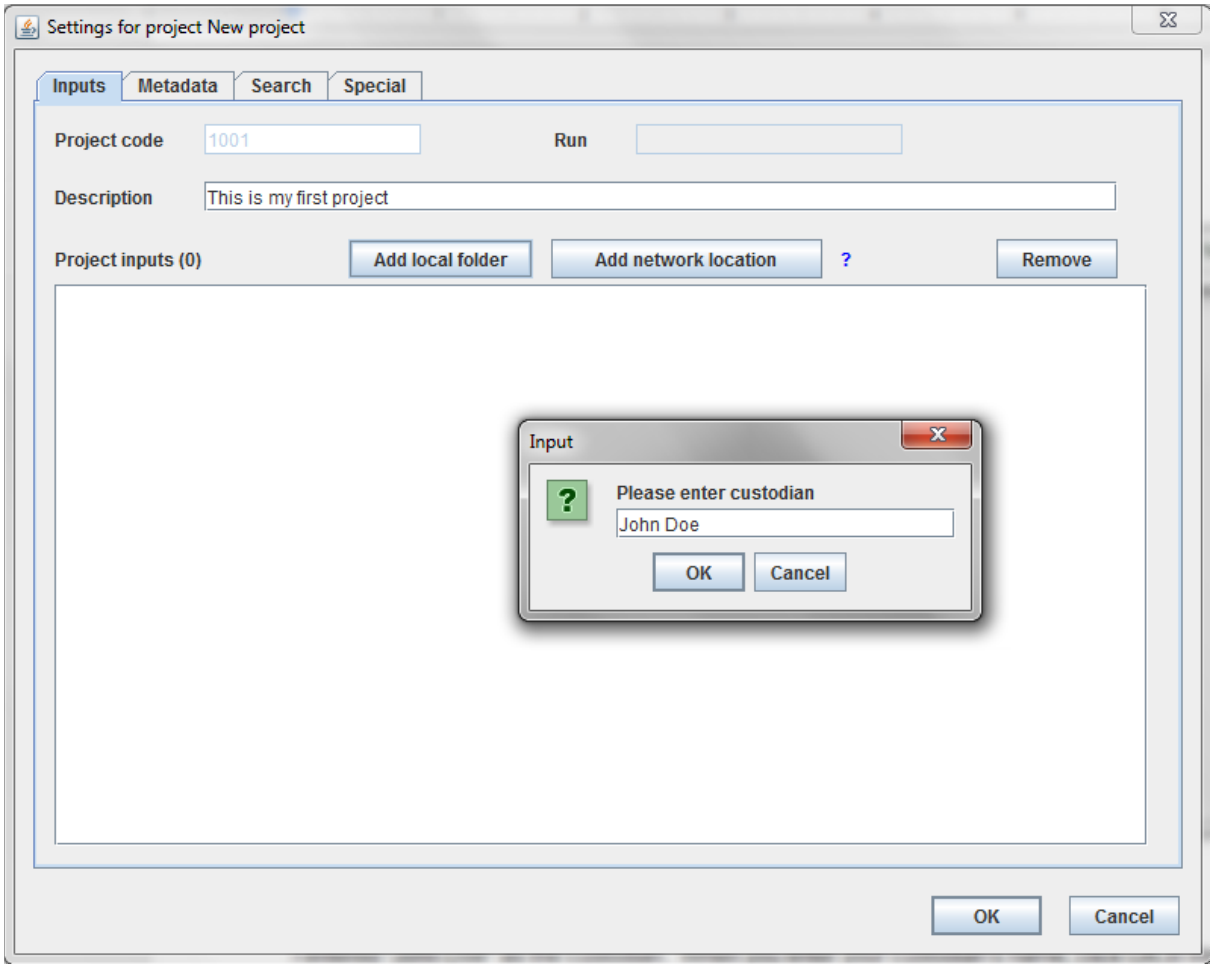


FIGURE 10.5

I entered "John Doe" as the custodian. When you enter your custodian's name, click OK in the Input window. The Input window will promptly close. This action will save your file path, as seen in Figure 10.6 below. Note that the name of the custodian is inserted at the beginning of the file path as shown in Figure 10.6.

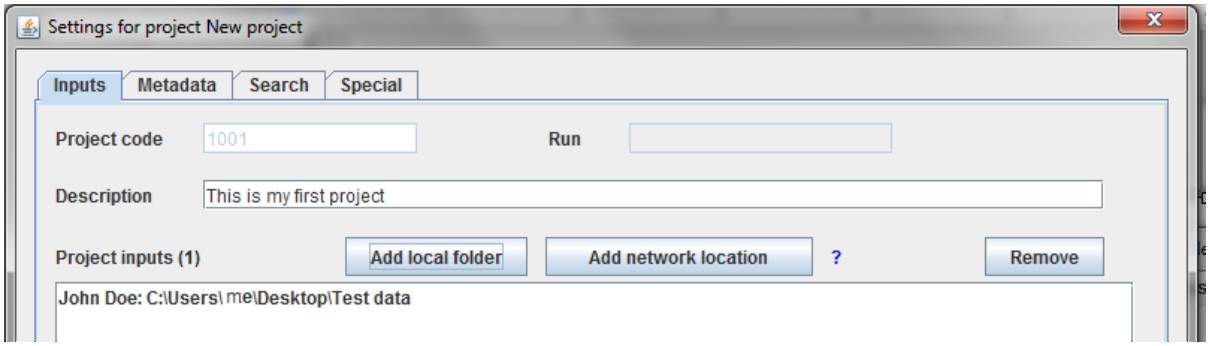


FIGURE 10.6

Clicking OK in the settings menu will bring up a Save screen, as seen in Figure 10.7. I've decided to save my project by the name "My Project-1". You can call your project by any name that you wish.

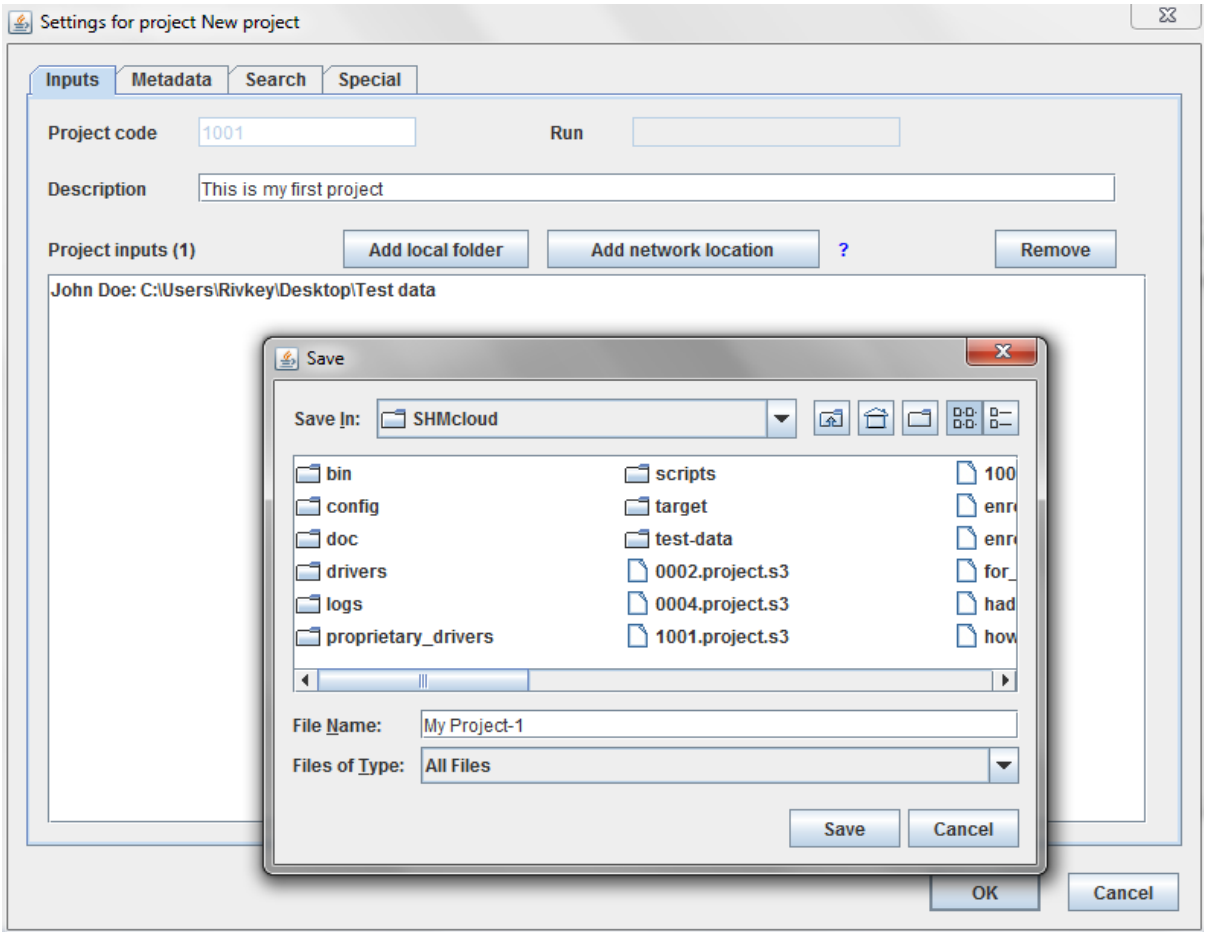


FIGURE 10.7

Notice the top of the Save screen where it says "Save In:". This of course refers to the directory where the project will be saved. In the future you can reopen this project by using the "Open" function and selecting this folder. I am saving my project in the default SHMcloud™ directory. You can choose to save your project wherever you want on your computer or external device.

Clicking "Save" in the Save screen will save your project and close both the "Save screen" and the "Settings for project screen".

Note: If you choose, you can also give your **File Name** the same name that you gave to your **Description**. This will not cause any conflict in processing. The Description is used internally once you have opened the project. The File name is the external name that your project is saved by on your computer. In other words, the File name is the file that you will choose to open in order to access the given project.

10.8 What if you saved your project, got interrupted, and came back and forgot what project you were working on, or what settings you put into place? The top of the SHMcloud™ menu still says “New project” because we have never reopened this project. No worries, you can easily check to see what project you currently have open.

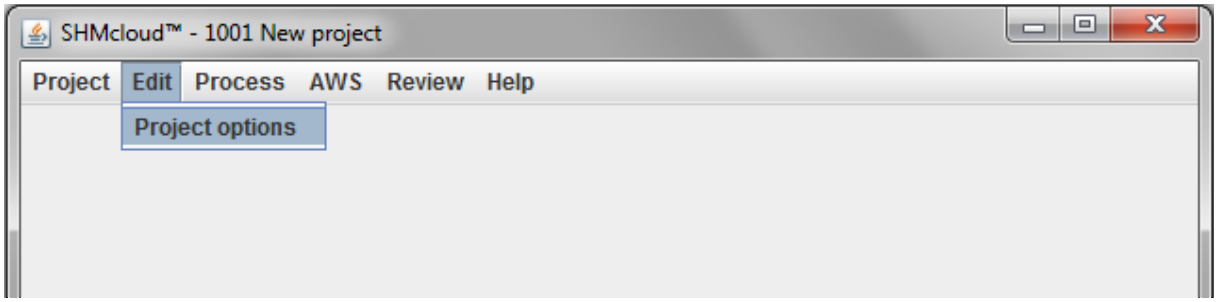


FIGURE 10.8

As seen in Figure 10.8 above, just click on “Edit” in your SHMcloud™ menu, and then “Project options”. The “Settings for project” screen will pop up (similar to Figure 10.6), and you will be able to see what project is currently open.

⇒ ***Please note,*** you may select more than one data set and assign the same or different Custodians to be run as part of a single project, simply by clicking “Add local folder” or “Add network location”, as we have done in the previous steps (Figures 10.3 through 10.6).

You can add as many additional folders and files to your project as you would like. Even though we have reopened this screen, we can still add more folders to process in the same run.

Each dataset will appear as a “Project Input” along with its path, as seen in Figure 10.6 above.

After processing your project, the output “Metafile” will define each Custodian accordingly.

10.9 *Question:* What if you added folders that you really do not want?

Simply highlight the undesired folders and click the “Remove” tab, as seen in Figure 10.9.

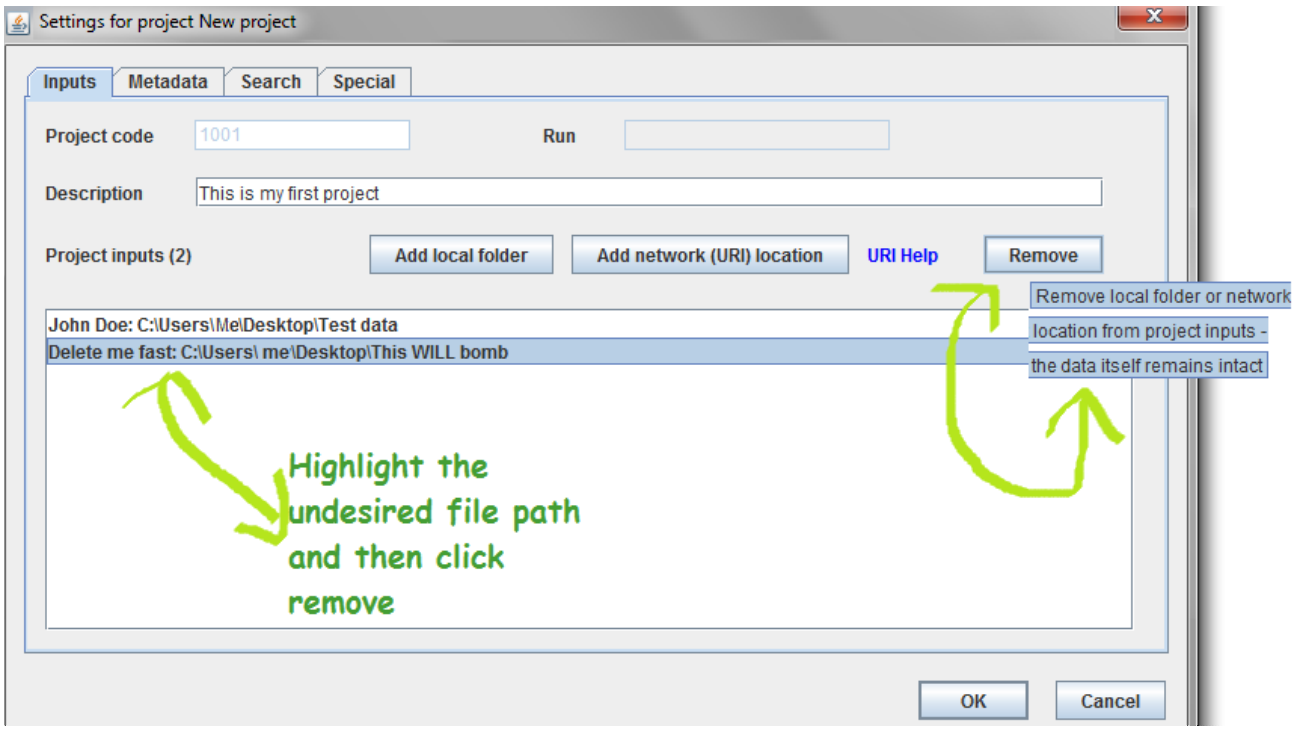


FIGURE 10.9

Clicking “OK” at the bottom of the “Settings” screen will save any changes that you may have made to your project.

⇒ By the way, did you know that each saved project has an internal code that identifies it? In the upper left corner of Figure 10.9 we see the Project code as being 1001. You do not need to remember this number, as you have given your project an identifiable Description and File Name. But you might want to be aware that a number will be created and will correspond uniquely to each individual project.

⇒ Did you notice that the top of the screen still refers to this project as “New Project”, even though we gave our project an identifiable description name? This is because we have not yet reopened our project.

Points to notice:

After you click OK in the Settings screen as seen in Figure 10.9, your project is saved. You are able to move onto Process and Stage your project at this point.

However, while we are here, let's notice a few other things about our screen. *(If you do not wish to notice anything, then feel free to skip down to 10.15.)*

When you look at your SHMcloud™ menu you will notice that the header still identifies your project as a "New Project". Do you see the number at the top of the screen? It is the same number that was listed as your Project code, as seen in Figure 10.9.

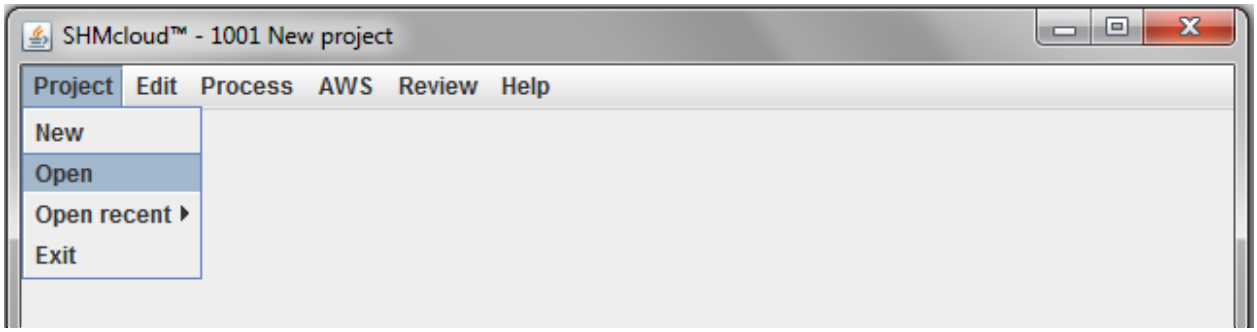


Figure 10.10

You may ask, what was the Description for? When will I ever see that? How about the File Name that I gave to my project? All I can see right now is an internal number that I have no control over.

To answer these questions, let us try to reopen an existing project. In this example we will open the project that we just created. Granted that project is already open, but our software will allow the user to reopen any saved project, including one that is currently open.

In your SHMcloud™ menu, click "Project" and then "Open", as seen in Figure 10.10.

The "Select project file" menu will open shown in Figure 10.11, below. Above, in Figure 10.7, I saved my project in the SHMcloud folder by the name My Project-1.

The file extension "project" was automatically given by the software.

I would like to reopen that project now. So I scrolled to "My Project-1.project", selected it, and clicked "Open".

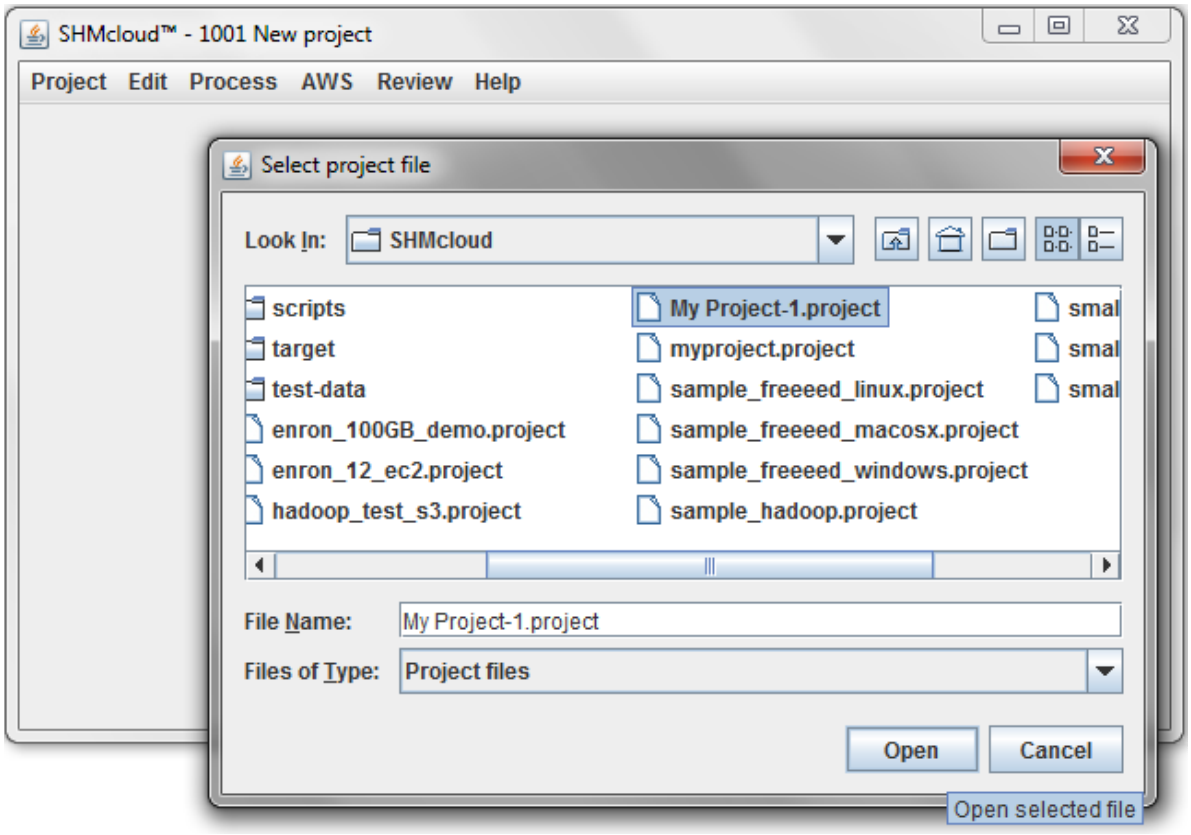


Figure 10.11

Immediately after we click to Open the project, the title at the top of our SHMcloud™ menu will change. The identifying number 1001 (specific to my project) still appears in the title at the top of this menu, however, the description name “The is my first project” also appears here.

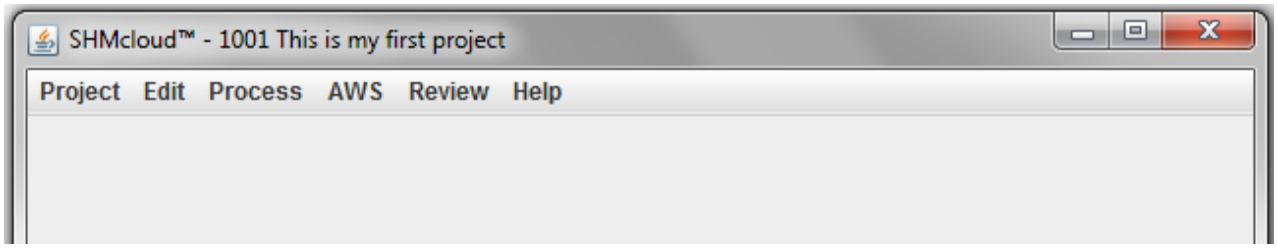


Figure 10.12

Since we never actually processed this project before, the Settings menu will once again open.

Notice the top of the Settings screen. The Description that I gave my project back in Figure 10.6 now also appears in the title at the top of this screen, Figure 10.13.

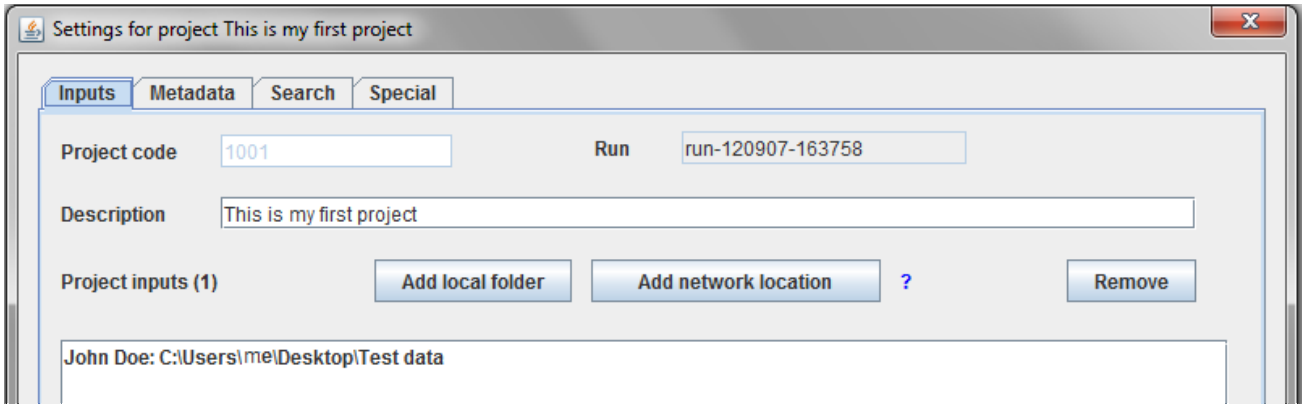


Figure 10.13

What if you already ran this project at least once before, and are now reopening it?

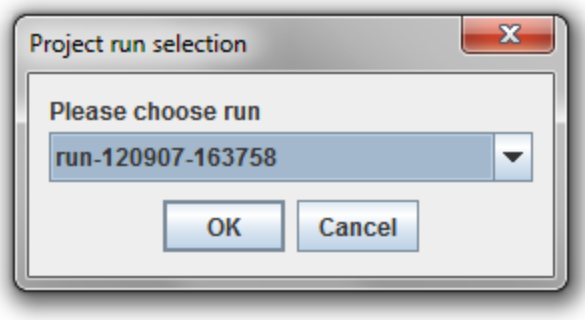


Figure 10.14

As mentioned earlier in Section 5, under such circumstances our software will give you a choice of which run you would like to open, Figure 10.14.

You get to choose.

See what your options are.

After you choose your project run, the settings menu will open, Figure 10.13.

10.15 Now we are ready to Process our project.

The steps moving forward will mirror of what we did earlier in Section 6, when we were checking the functionality of our program using sample data.

⇒ We have just created and saved our project with the project files specified.

⇒ We are now ready to **"Stage"** the data. In a nutshell, Staging zips up the data in preparation for processing.

As mentioned earlier, it is important to note that Staging must be done before any project can be run, regardless of whether it is Processed Locally or run in the cloud using AWS.

First we will Stage the new Project as shown in Figure 10.15. We initially discussed Staging earlier in Section 6.

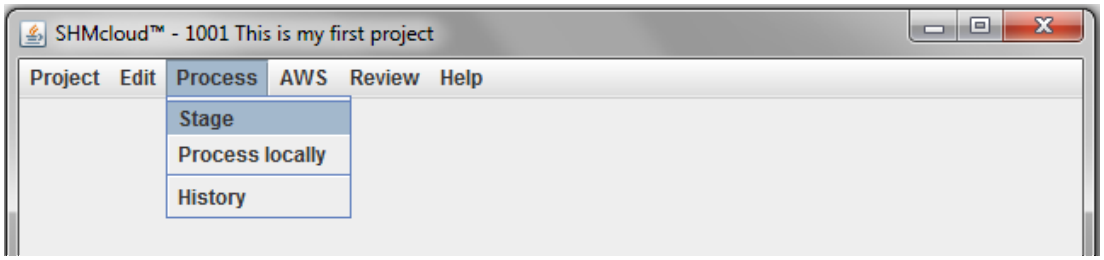


FIGURE 10.15

Note that if we would have continued from the beginning with a “New project”, then the Title bar would still be displaying “New project” in the title at the top of our menu, along with the identifying project number. As stated before, the title bar will not reflect the identifying description name until you reopen the project.

Also note that the output messages from **<u>Staging</u>** will appear in the *“**command window**”* and/or in the ***“processing history window”,*** as seen below in Figure 10.16.

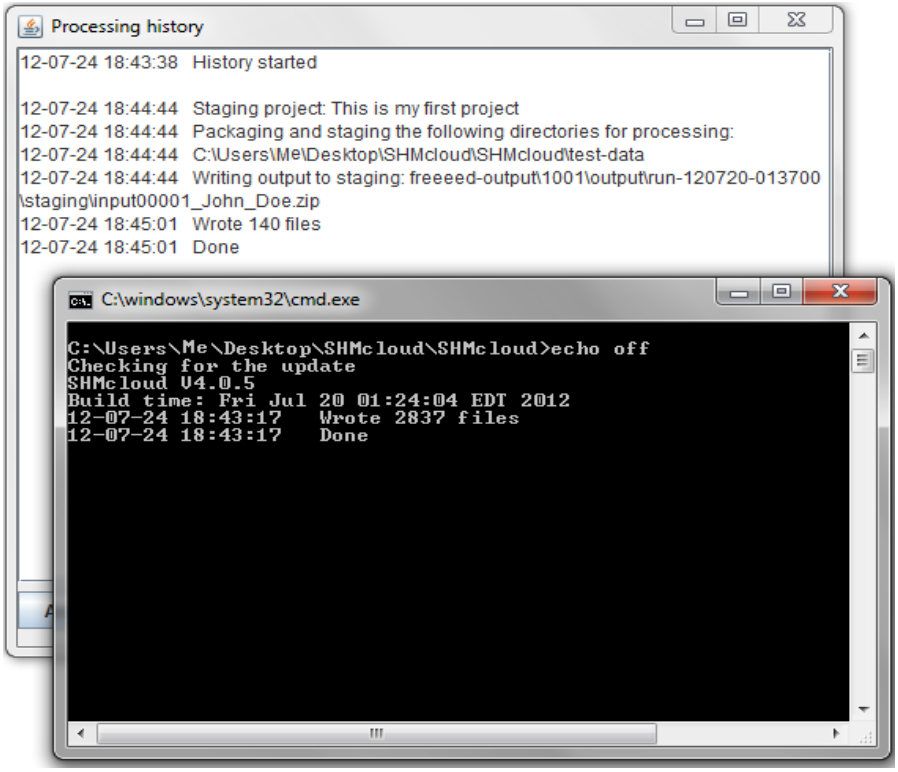

Processing history
12-07-24 18:43:38 History started
12-07-24 18:44:44 Staging project: This is my first project
12-07-24 18:44:44 Packaging and staging the following directories for processing:
12-07-24 18:44:44 C:\Users\Me\Desktop\SHMcloud\SHMcloud\test-data
12-07-24 18:44:44 Writing output to staging: freeeed-output\1001\output\run-120720-013700
\staging\input00001_John_Doe.zip
12-07-24 18:45:01 Wrote 140 files
12-07-24 18:45:01 Done
C:\windows\system32\cmd.exe
C:\Users\Me\Desktop\SHMcloud\SHMcloud>echo off
Checking for the update
SHMcloud V4.0.5
Build time: Fri Jul 20 01:24:04 EDT 2012
12-07-24 18:43:17 Wrote 2837 files
12-07-24 18:43:17 Done

FIGURE 10.16

When the output message indicates this step is done (as seen in Figure 10.16), then you are ready to set up your **Amazon environment** and begin processing your staged data. ***We will discuss how to set up an AWS environment in Section 11.*** Meanwhile, assuming our test data is small enough, we will process our data using the free local processor that SHMcloud™ provides.

Select Process from the Process pull down menu as shown in Figure 10.17.

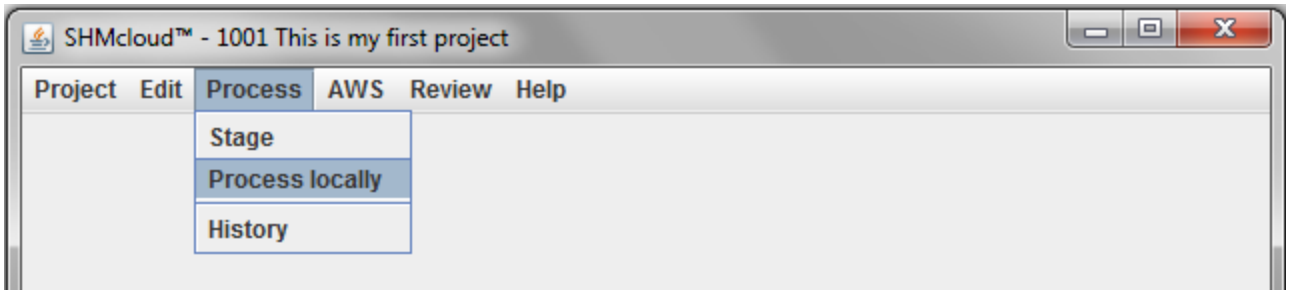


FIGURE 10.17

While you are processing, your **Processing history** screen and your **CMD** screen will be very busy. You will know that your job is finished when you see the word "Done" appear at the bottom of the screens, as seen in Figure 10.18.

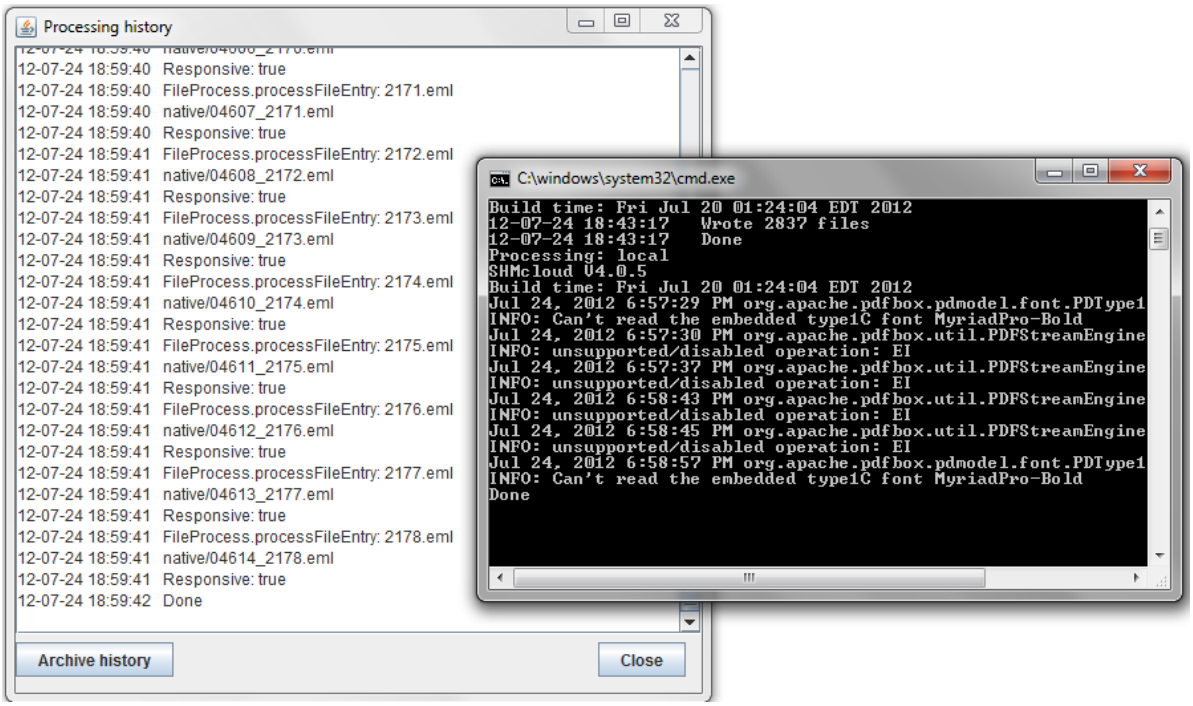

Processing history
12-07-24 18:59:40 Responsive: true
12-07-24 18:59:40 FileProcess.processFileEntry: 2171.eml
12-07-24 18:59:40 native/04607_2171.eml
12-07-24 18:59:40 Responsive: true
12-07-24 18:59:41 FileProcess.processFileEntry: 2172.eml
12-07-24 18:59:41 native/04608_2172.eml
12-07-24 18:59:41 Responsive: true
12-07-24 18:59:41 FileProcess.processFileEntry: 2173.eml
12-07-24 18:59:41 native/04609_2173.eml
12-07-24 18:59:41 Responsive: true
12-07-24 18:59:41 FileProcess.processFileEntry: 2174.eml
12-07-24 18:59:41 native/04610_2174.eml
12-07-24 18:59:41 Responsive: true
12-07-24 18:59:41 FileProcess.processFileEntry: 2175.eml
12-07-24 18:59:41 native/04611_2175.eml
12-07-24 18:59:41 Responsive: true
12-07-24 18:59:41 FileProcess.processFileEntry: 2176.eml
12-07-24 18:59:41 native/04612_2176.eml
12-07-24 18:59:41 Responsive: true
12-07-24 18:59:41 FileProcess.processFileEntry: 2177.eml
12-07-24 18:59:41 native/04613_2177.eml
12-07-24 18:59:41 Responsive: true
12-07-24 18:59:41 FileProcess.processFileEntry: 2178.eml
12-07-24 18:59:41 native/04614_2178.eml
12-07-24 18:59:41 Responsive: true
12-07-24 18:59:42 Done
Archive history
Close
C:\windows\system32\cmd.exe
Build time: Fri Jul 20 01:24:04 EDT 2012
12-07-24 18:43:17 Wrote 2837 files
12-07-24 18:43:17 Done
Processing: local
SHMcloud V4.0.5
Build time: Fri Jul 20 01:24:04 EDT 2012
Jul 24, 2012 6:57:29 PM org.apache.pdfbox.pdmodel.font.PDType1
INFO: Can't read the embedded type1C font MyriadPro-Bold
Jul 24, 2012 6:57:30 PM org.apache.pdfbox.util.PDFStreamEngine
INFO: unsupported/disabled operation: EI
Jul 24, 2012 6:57:37 PM org.apache.pdfbox.util.PDFStreamEngine
INFO: unsupported/disabled operation: EI
Jul 24, 2012 6:58:43 PM org.apache.pdfbox.util.PDFStreamEngine
INFO: unsupported/disabled operation: EI
Jul 24, 2012 6:58:45 PM org.apache.pdfbox.util.PDFStreamEngine
INFO: unsupported/disabled operation: EI
Jul 24, 2012 6:58:57 PM org.apache.pdfbox.pdmodel.font.PDType1
INFO: Can't read the embedded type1C font MyriadPro-Bold
Done

FIGURE 10.18

At this point we have processed our own project locally. We are now ready to go to the Review Menu and pull down “Open output folder” and view your results, just as we have done previously in section #9.

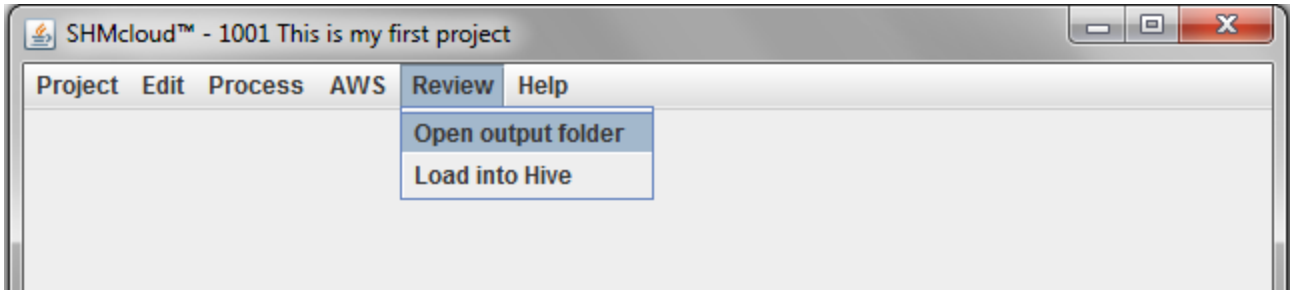


FIGURE 10.19

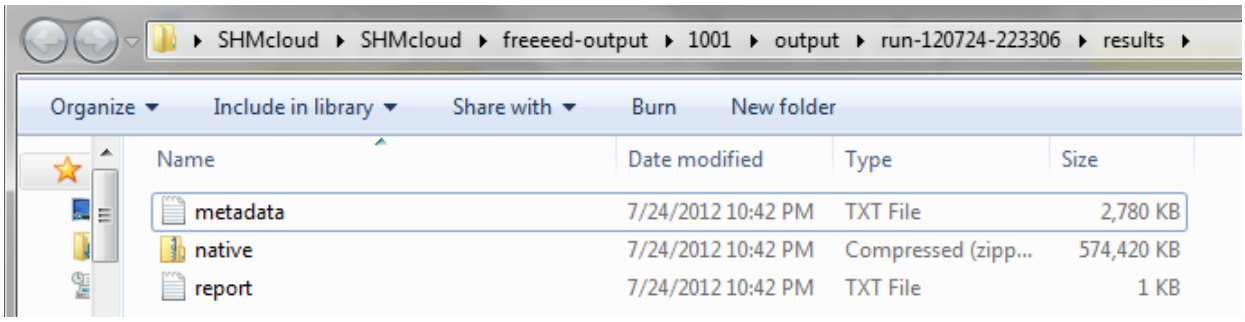


FIGURE 10.20

Notes and Warnings:

Now that we understand how to access our project output, let’s discuss it in a bit more detail.

When you click on **Review** then **Open output folder**, as seen in Figure 10.19, you will see three files/folders in the review folder, Figure 10.20.

If there are only two files in your output folder, then chances are you are missing the file called “**report**”. The report file will only appear when your project has finished running. Do not attempt to open your metadata file until after the project has finished running!

⇛ ***metadata*** *- is your project output load file, as discussed in detail in Section 9. This is the output that you are looking for when you run your project.*

⇛ ***native*** *- is a zipped folder. It contains all extracted native files, including emails and text extracted from them, as well as "exception" files that could not get processed for any reason. Essentially it is everything that this project processed.*

⇛ ***report*** *- is a simple report of your run. It contains the name of your project, when it started, when it finished, how long it took to run, and how many items were included in this run.*

That explains the basics of your output folder.

Now that you know where to find your output, what happens if you decide to sneak a peek at it while the project is still running?

WARNING!! DO NOT DO THAT!!

If you try to open your **metafile** while the SHMcloud™ Player is still processing, it will cease to continue. Yes, the **metafile** will actually open, but it will also no longer be written in.

Your output will be incomplete.

There will be no warning from the Player, and nothing will stop you from doing it. So consider this to be your only warning!

Additionally, if you open your **metafile** while your project is still running, your **results folder** will not produce the **report** file that we discussed above. Perhaps the lack of a final report on the project will be a sign for you to realize that you interrupted the project mid-run.

By the way, if you are running a project and you are waiting to see when it will be completed, you can keep your output folder open. As long as only two files appear there, you will know that your project is still running.

When your project has completed running, a third file will appear. But instead of being called "Report" as we just mentioned, the file will show up as "SUCCESS", as seen in Figure 10.21.

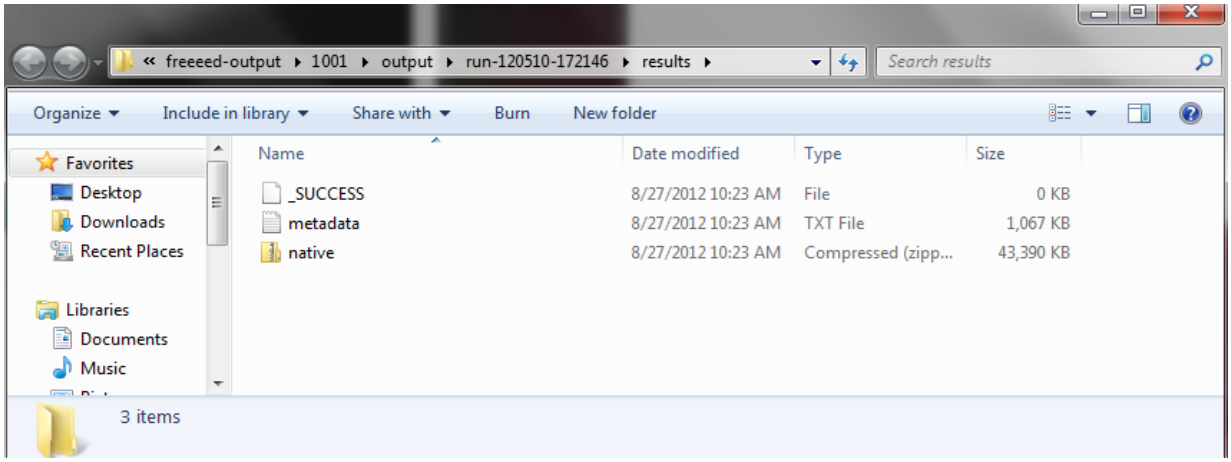


Figure 10.21

Of course this means that your project was successfully run! If you click on the SUCCESS file, you will probably not be able to open it. Reopening the output folder will cause the "SUCCESS" line to disappear, and "report" to appear instead.

The trick to keeping your output folder open while you are processing is so you will know that when SUCCESS appears, your metadata file is ready to be opened.

Multiple Output Files

Another interesting point to note is that for each Project that you run, you can go to the Review menu for that given project and view the output.

⇛ Just open a previously run project, go to the **Review** menu, click on "**open output folder**", and viola! your project output for that particular run is still saved and ready for you to see.

⇛ Your Review files will only get overwritten if you rerun the same project.

⇛ This means that if you have several different projects (each with a different name) you will also have multiple output folders.

⇛ Of course if you want to guarantee that you do not lose any of your output, copy it from your SHMcloud™ Player and save it with a distinctive name someplace else on your computer or external storage device.

Later, in Section 15, we will discuss how to process your projects using specific search options. But for now we will first go through the basic steps of processing projects in the cloud.

11. Setting up an Amazon AWS Account

Before you can actually process any of your projects in the cloud, you will need to have an Amazon Web Service (AWS) account.

If you already have an Amazon Web Service (AWS) account, then you may skip this section and continue with section #12.

Setting up an Amazon Web Service (AWS) account is free and easy, and you only pay for what you use in storage and processing time. The processing and storage capacity are unlimited, so you can use as much or as little as you need -- and only pay for what you use.

You will have access to storage with Amazon S3, and computing resources with Amazon's Elastic Compute (EC2) environment, as well as many other resources from Amazon.

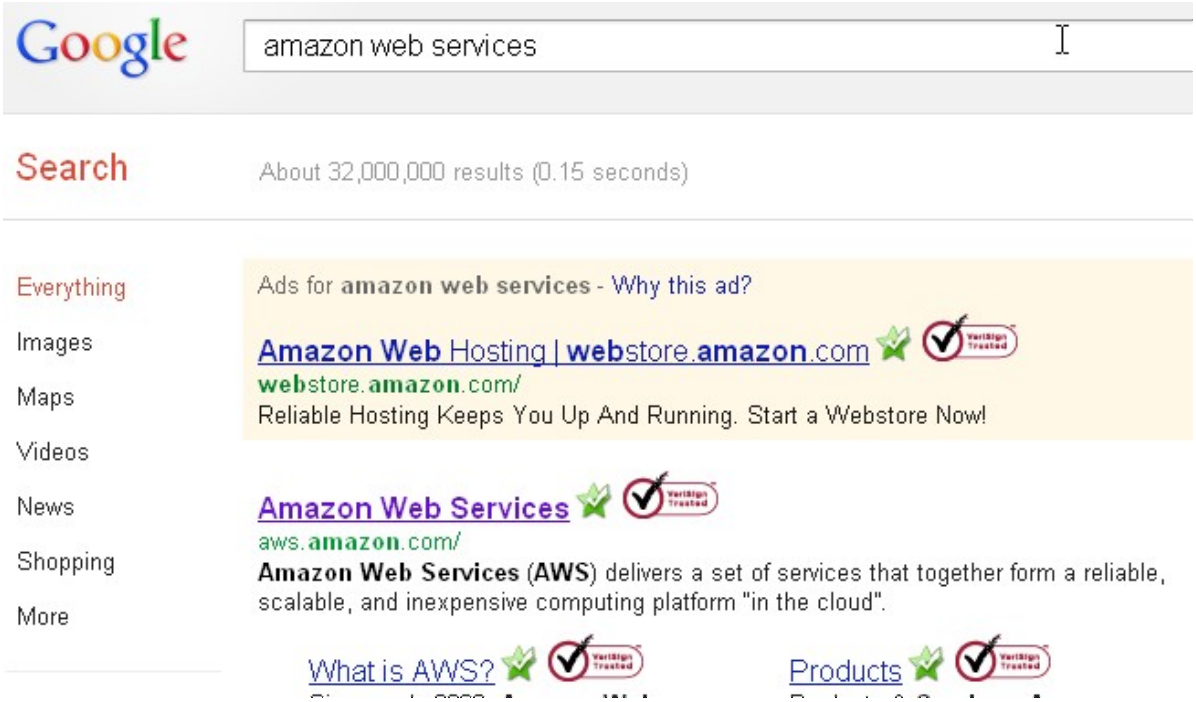
The account setup takes just a few minutes and entails the following steps:

1. In your web browser search for "Amazon Web Services", or just go to http://aws.amazon.com/
2. Choose to sign up, and enter your email address and a password for your AWS account
3. Confirm your name, email address and password
4. Provide your contact information (address and phone number)
5. Read and agree to the terms of service
6. Provide your payment information (credit card, but no charges will be made yet)
7. Confirm your phone number (automated call to your number you provided)
8. Receive confirmation screen and email that your account is active

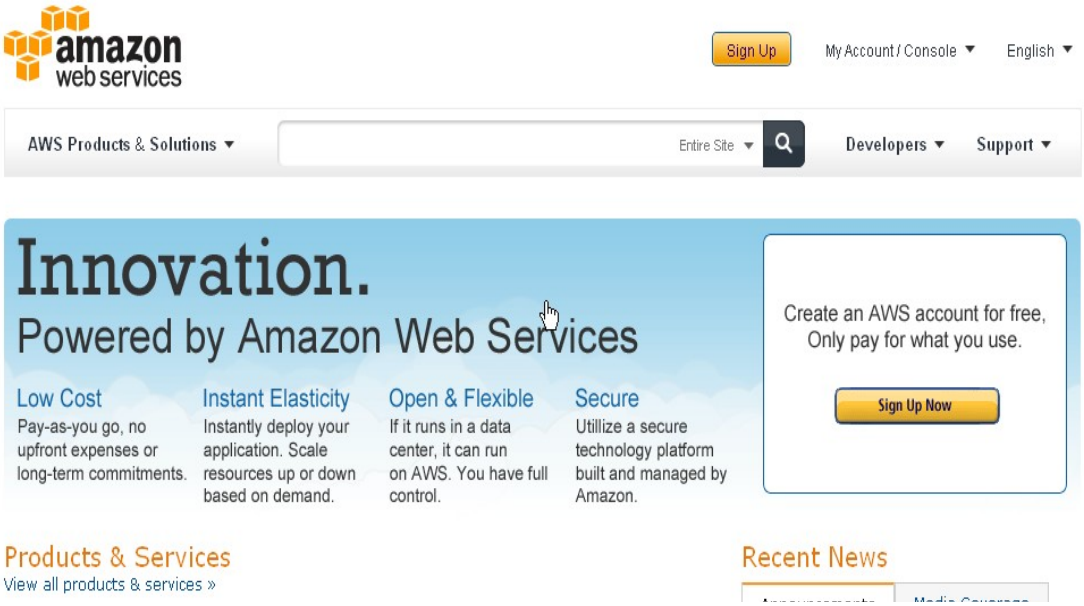
That's it!

Below are examples of these simple steps, with screenshots included.

11.1. In your web browser, search for “Amazon Web Services”, or just go to http://aws.amazon.com/.



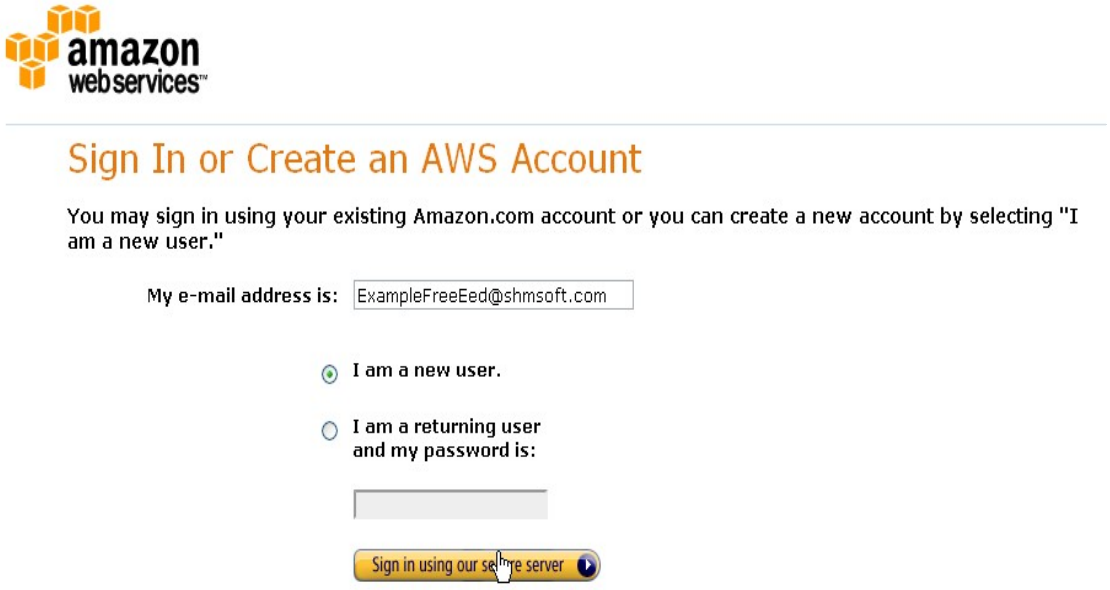
Below is what the home page for Amazon Web Services (aws.amazon.com) looks like:


amazon web services
Sign Up
My Account / Console
English
AWS Products & Solutions
Entire Site
Developers
Support
Innovation.
Powered by Amazon Web Services
Low Cost
Pay-as-you go, no upfront expenses or long-term commitments.
Instant Elasticity
Instantly deploy your application. Scale resources up or down based on demand.
Open & Flexible
If it runs in a data center, it can run on AWS. You have full control.
Secure
Utilize a secure technology platform built and managed by Amazon.
Create an AWS account for free, Only pay for what you use.
Sign Up Now
Products & Services
View all products & services »
Recent News

11.2. Choose to sign up, and enter your email address and a password for your AWS account.

Click the “Sign Up” button, and you will see a page like this:

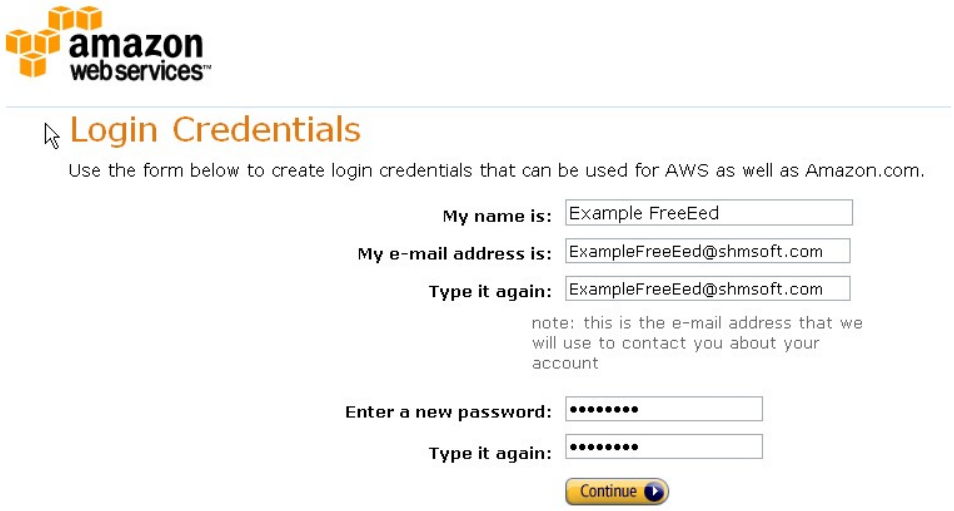
You can also get to the following Amazon signon page by linking here:
https://portal.aws.amazon.com/gp/aws/user/subscription/index.html?offeringCode=14A5AD2D


amazon
web services
Sign In or Create an AWS Account
You may sign in using your existing Amazon.com account or you can create a new account by selecting "I am a new user."
My e-mail address is:
ExampleFreeEed@shmsoft.com
I am a new user.
I am a returning user and my password is:
Sign in using our secure server

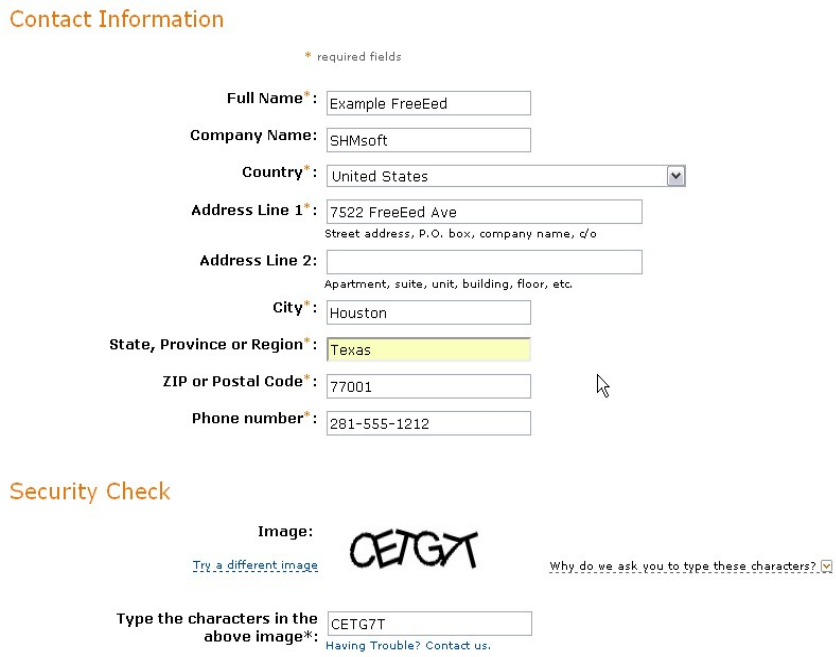
Enter your email address, and click the “Sign in...” button

11.3. Confirm your name, email address and password.

This is the email and password you will use from now on to logon onto your AWS account. Choose a password that is secure that you will remember. This may be the same as the password you use for your email, but it can be different if you would like.


amazon web services
Login Credentials
Use the form below to create login credentials that can be used for AWS as well as Amazon.com.
My name is: Example FreeEed
My e-mail address is: ExampleFreeEed@shmsoft.com
Type it again: ExampleFreeEed@shmsoft.com
note: this is the e-mail address that we will use to contact you about your account
Enter a new password: ••••••••
Type it again: ••••••••
Continue

11.4. Provide your contact information (address and phone number).


Contact Information
* required fields
Full Name*: Example FreeEed
Company Name: SHMsoft
Country*: United States
Address Line 1*: 7522 FreeEed Ave
Street address, P.O. box, company name, c/o
Address Line 2:
Apartment, suite, unit, building, floor, etc.
City*: Houston
State, Province or Region*: Texas
ZIP or Postal Code*: 77001
Phone number*: 281-555-1212
Security Check
Image:
Try a different image
CETG7T
Why do we ask you to type these characters?
Type the characters in the above image*: CETG7T
Having Trouble? Contact us.

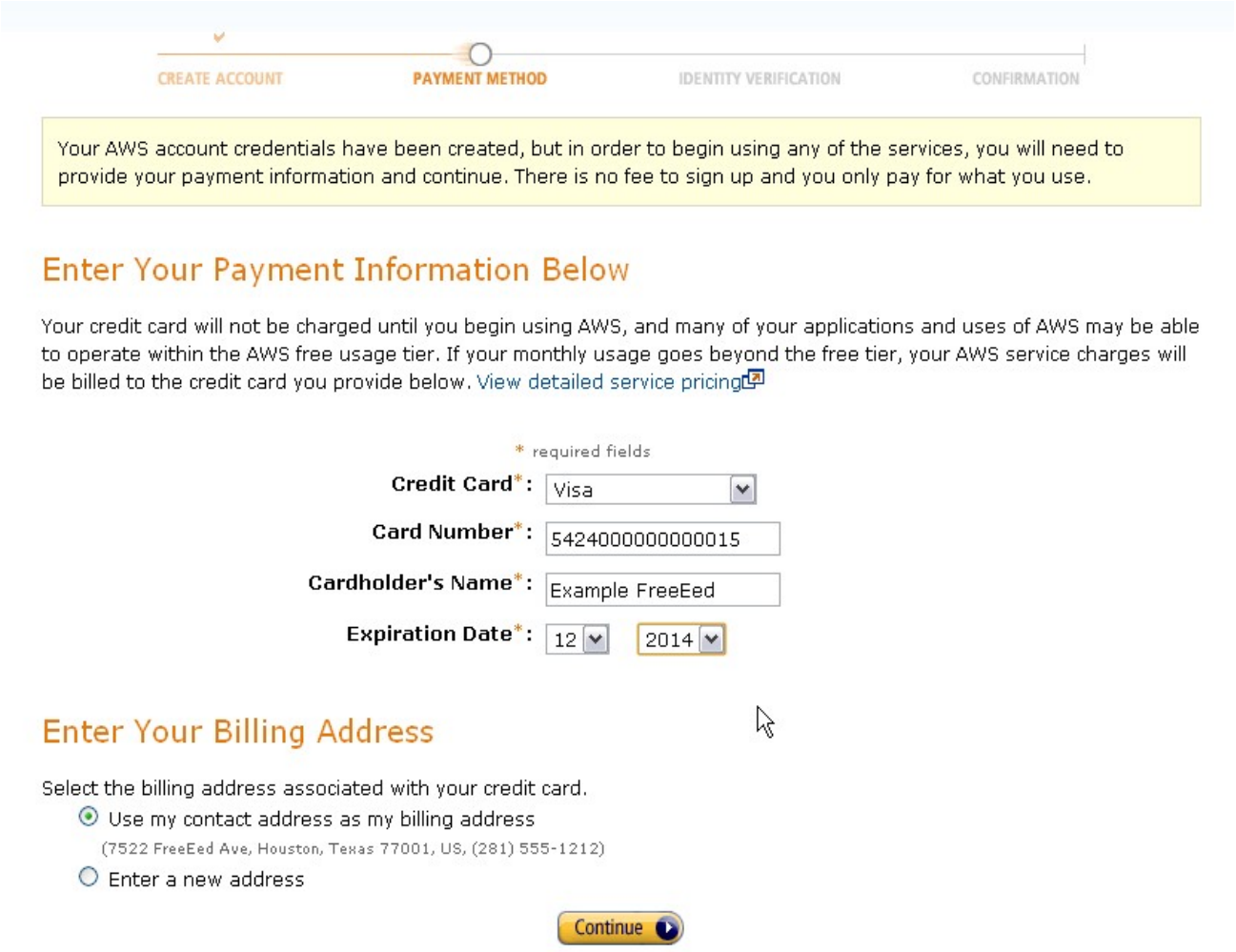
You will need to enter the scrambled characters as well, to confirm that you are a person signing up for an account, rather than some automated process.

11.5. Read and agree to the terms of service

AWS Customer Agreement
Check here to indicate that you have read and agree to the terms of the
Amazon Web Services Customer Agreement.
Create Account and Continue

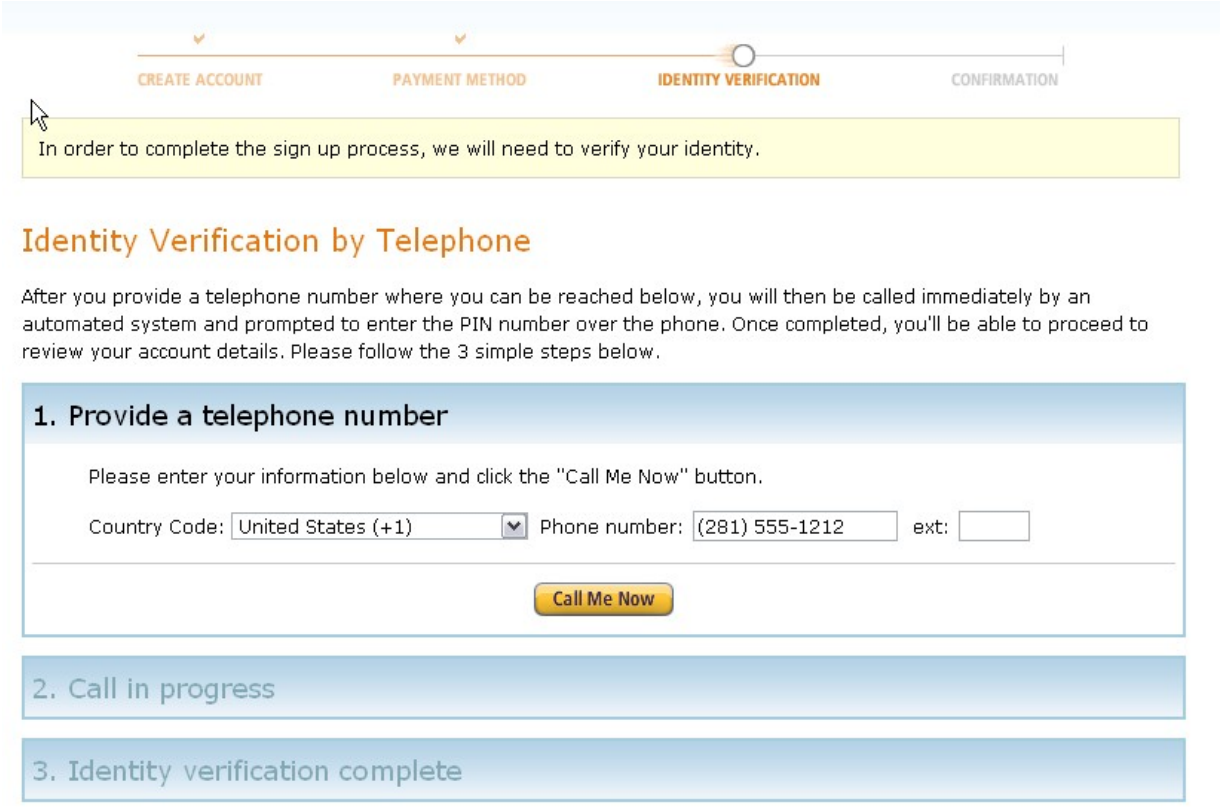
11.6. Provide your payment information (credit card, but no charges will be made yet).

Note: You will only start accruing charges for your projects when you click on "Start cluster" (explained in a later section) and never before.


CREATE ACCOUNT
PAYMENT METHOD
IDENTITY VERIFICATION
CONFIRMATION
Your AWS account credentials have been created, but in order to begin using any of the services, you will need to provide your payment information and continue. There is no fee to sign up and you only pay for what you use.
Enter Your Payment Information Below
Your credit card will not be charged until you begin using AWS, and many of your applications and uses of AWS may be able to operate within the AWS free usage tier. If your monthly usage goes beyond the free tier, your AWS service charges will be billed to the credit card you provide below. View detailed service pricing
* required fields
Credit Card*: Visa
Card Number*: 5424000000000015
Cardholder's Name*: Example FreeEed
Expiration Date*: 12 2014
Enter Your Billing Address
Select the billing address associated with your credit card.
Use my contact address as my billing address
(7522 FreeEed Ave, Houston, Texas 77001, US, (281) 555-1212)
Enter a new address
Continue

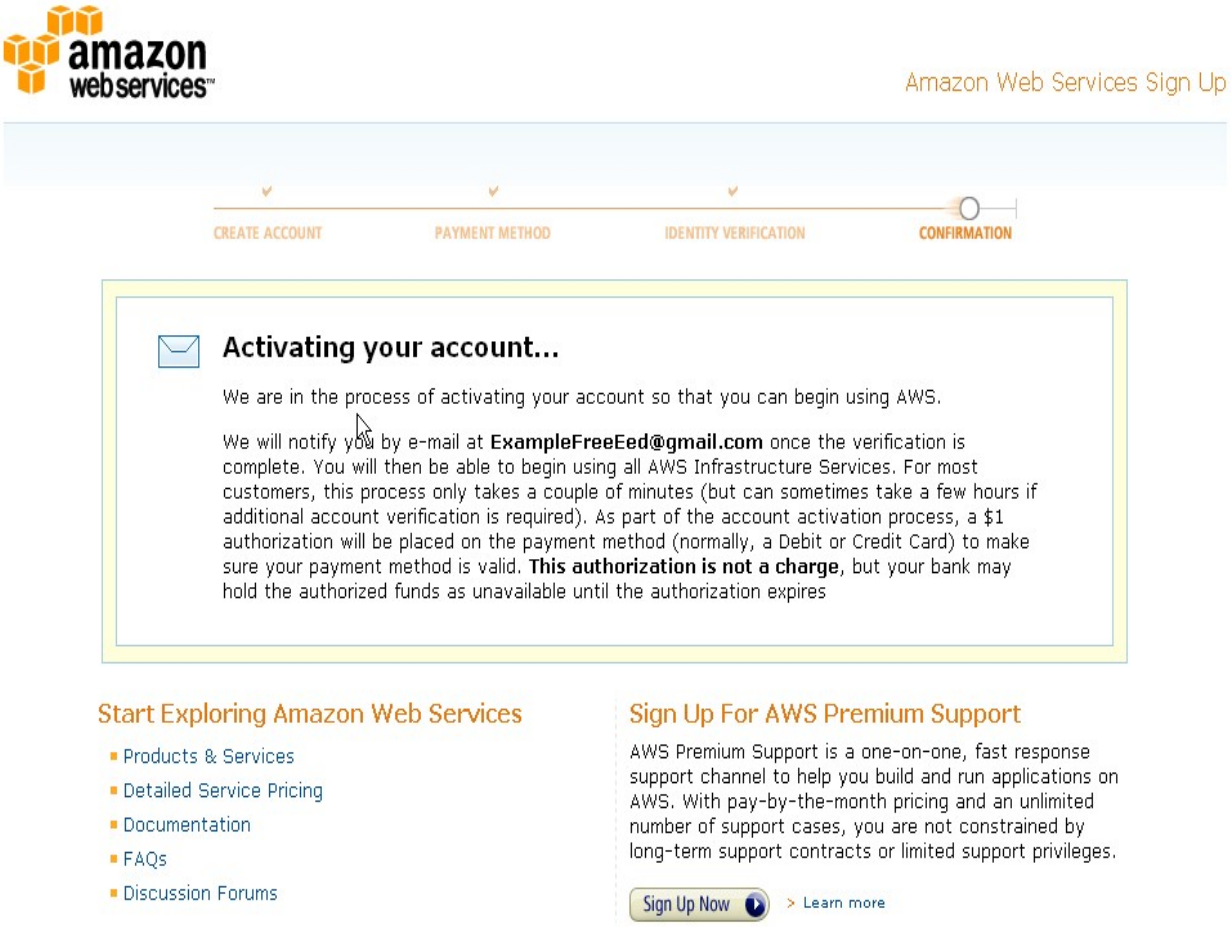
11.7. Confirm your phone number (automated call to your number you provided)

For this step, the amazon web page will provide a confirmation code, a PIN number.
Then an automated call will be made to the phone number you provide.
You answer and listen to a recording asking you to enter the PIN provided.
You enter the PIN and now the phone number has been confirmed.


CREATE ACCOUNT
PAYMENT METHOD
IDENTITY VERIFICATION
CONFIRMATION
In order to complete the sign up process, we will need to verify your identity.
Identity Verification by Telephone
After you provide a telephone number where you can be reached below, you will then be called immediately by an automated system and prompted to enter the PIN number over the phone. Once completed, you'll be able to proceed to review your account details. Please follow the 3 simple steps below.
1. Provide a telephone number
Please enter your information below and click the "Call Me Now" button.
Country Code: United States (+1)
Phone number: (281) 555-1212
ext:
Call Me Now
2. Call in progress
3. Identity verification complete

11.8. Receive confirmation screen and email that your account is active.

You will see a confirmation page, and you will receive a confirmation email.


amazon web services
Amazon Web Services Sign Up
CREATE ACCOUNT
PAYMENT METHOD
IDENTITY VERIFICATION
CONFIRMATION
Activating your account...
We are in the process of activating your account so that you can begin using AWS.
We will notify you by e-mail at ExampleFreeEed@gmail.com once the verification is complete. You will then be able to begin using all AWS Infrastructure Services. For most customers, this process only takes a couple of minutes (but can sometimes take a few hours if additional account verification is required). As part of the account activation process, a $1 authorization will be placed on the payment method (normally, a Debit or Credit Card) to make sure your payment method is valid. This authorization is not a charge, but your bank may hold the authorized funds as unavailable until the authorization expires
Start Exploring Amazon Web Services
Products & Services
Detailed Service Pricing
Documentation
FAQs
Discussion Forums
Sign Up For AWS Premium Support
AWS Premium Support is a one-on-one, fast response support channel to help you build and run applications on AWS. With pay-by-the-month pricing and an unlimited number of support cases, you are not constrained by long-term support contracts or limited support privileges.
Sign Up Now
> Learn more

That's it! Your AWS account is now ready to use.

We don't use the application key at the moment, so once you register, you are done!

12. Processing your project in the Cloud

12.1 - In section #11 we walked through the steps of how to set up an Amazon Web Service (AWS) account.

Now we are ready to set up access to our Amazon environment, including S3 *(Simple Storage Service)* and EC2 *(Elastic Compute Cloud)*.

Of course you must already have a project set for processing. Follow through steps 10.1 through 10.16 from above until you have completed the Staging process for your project. Once your project has been properly "Staged", you are ready to process. However we will not be "Processing Locally", as we did in Section 10. Instead we would like to process in the Cloud, and so we continue from this point.

We want to use the supercomputers that we have available to us on Amazon when we have large amounts of Big Data for processing. But for the purpose of learning how to use AWS, we will continue these examples with our "Test data".

Go back to the SHMcloud™ Player. Notice the AWS menu, as seen in figure 12.1.

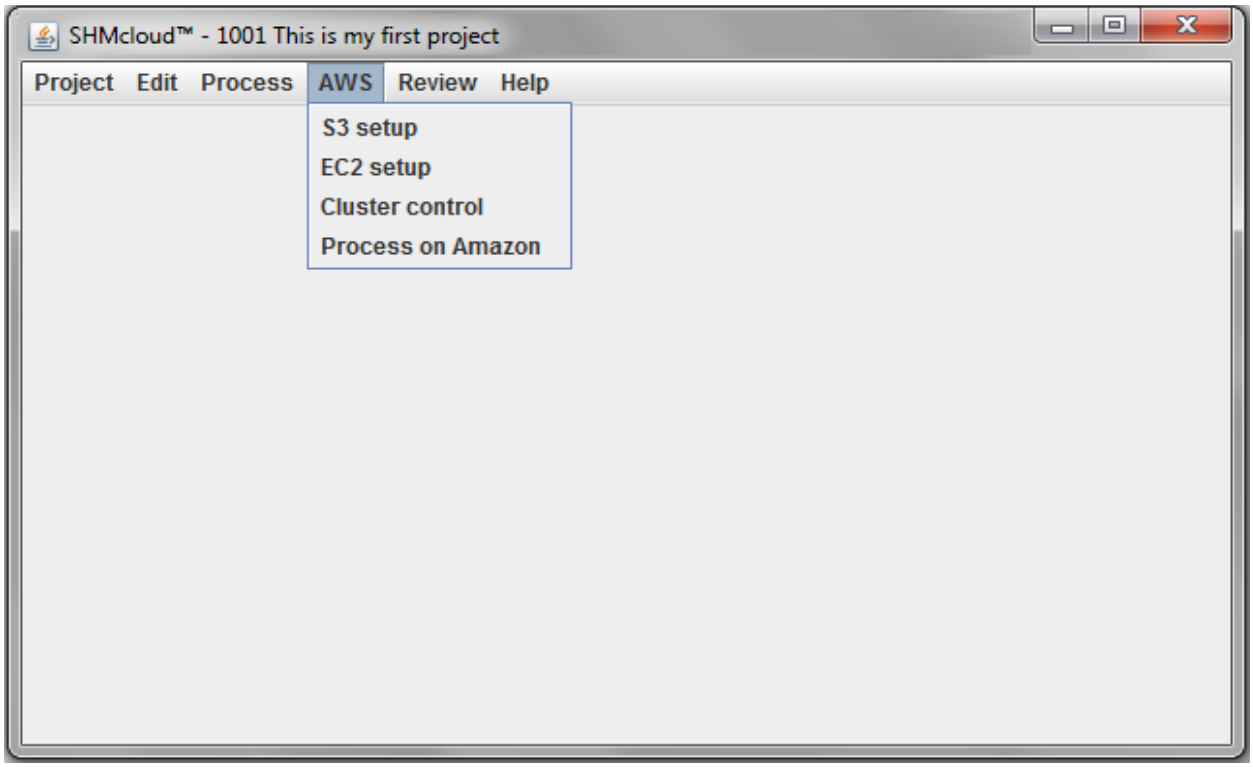


Figure 12.1

12.2 - Soon (as seen in Figure 12.5) you will be asked to provide the S3 keys which are in your Amazon account.

If you already know how to find your Amazon S3 keys, then you can skip down to 12.4. Otherwise, please keep reading.

To find your Amazon S3 keys:

Log in to your account at www.aws.amazon.com and choose “Security Credentials” from the menu. NOTE: If you are already logged into the AWS console, then choose “Security Credentials” from your account menu in the upper right of the page.

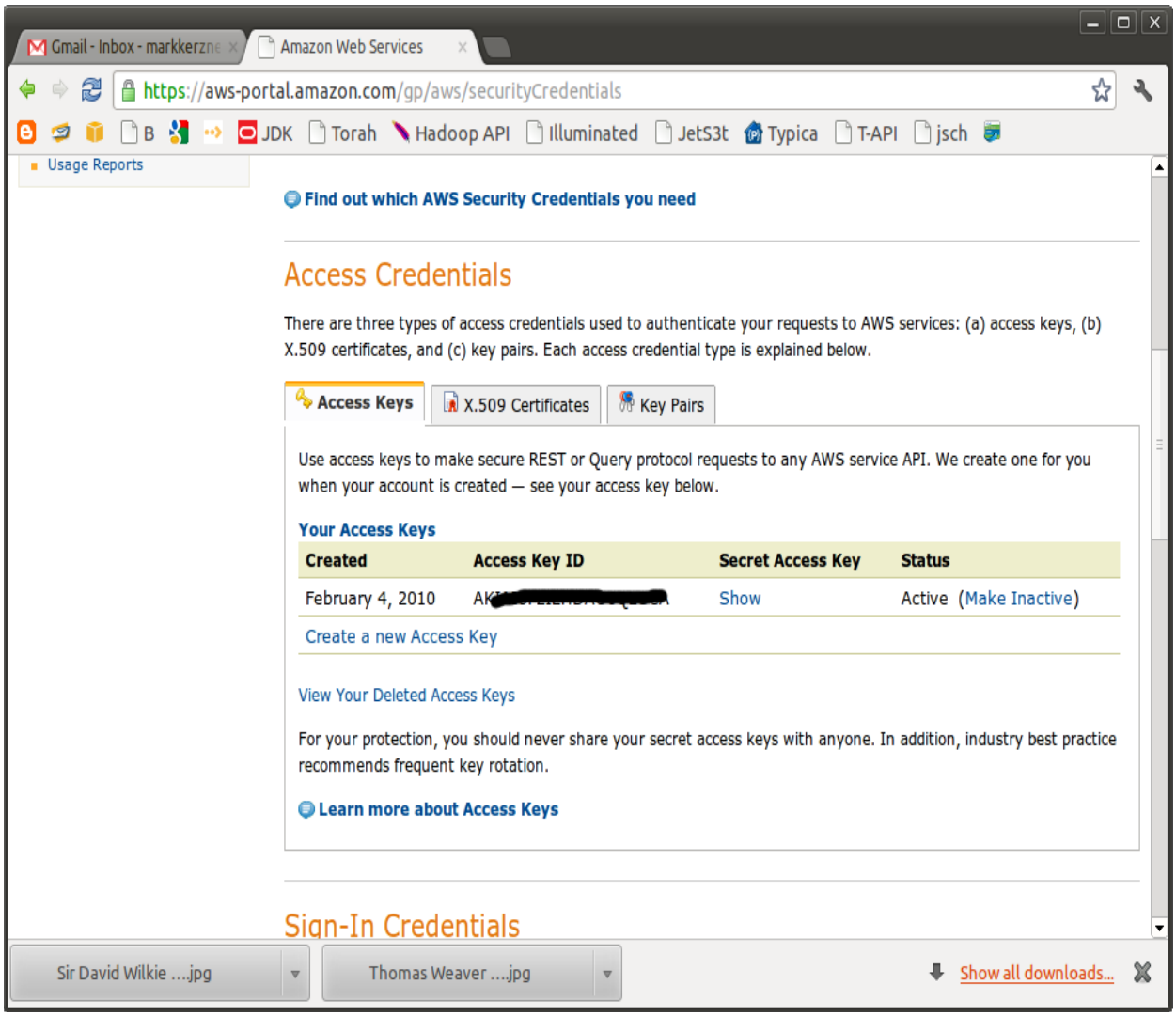

Gmail - Inbox - markkerzne
Amazon Web Services
https://aws-portal.amazon.com/gp/aws/securityCredentials
JDK
Torah
Hadoop API
Illuminated
JetS3t
Typica
T-API
jsch
Usage Reports
Find out which AWS Security Credentials you need
Access Credentials
There are three types of access credentials used to authenticate your requests to AWS services: (a) access keys, (b) X.509 certificates, and (c) key pairs. Each access credential type is explained below.
Access Keys
X.509 Certificates
Key Pairs
Use access keys to make secure REST or Query protocol requests to any AWS service API. We create one for you when your account is created — see your access key below.
Your Access Keys
Created
Access Key ID
Secret Access Key
Status
February 4, 2010
Show
Active (Make Inactive)
Create a new Access Key
View Your Deleted Access Keys
For your protection, you should never share your secret access keys with anyone. In addition, industry best practice recommends frequent key rotation.
Learn more about Access Keys
Sign-In Credentials
Sir David Wilkie ....jpg
Thomas Weaver ....jpg
Show all downloads...

Figure 12.2

12.3 - You will need to copy the Access Key ID and Secret Key ID to the corresponding fields of the SHMcloud™ setup (which we will soon see in Figure 12.5). Below we have blanked out the Access Key ID and the Secret Access Key (for our own security!). You will need to copy and paste those keys from your own account into the S3 setup screen.

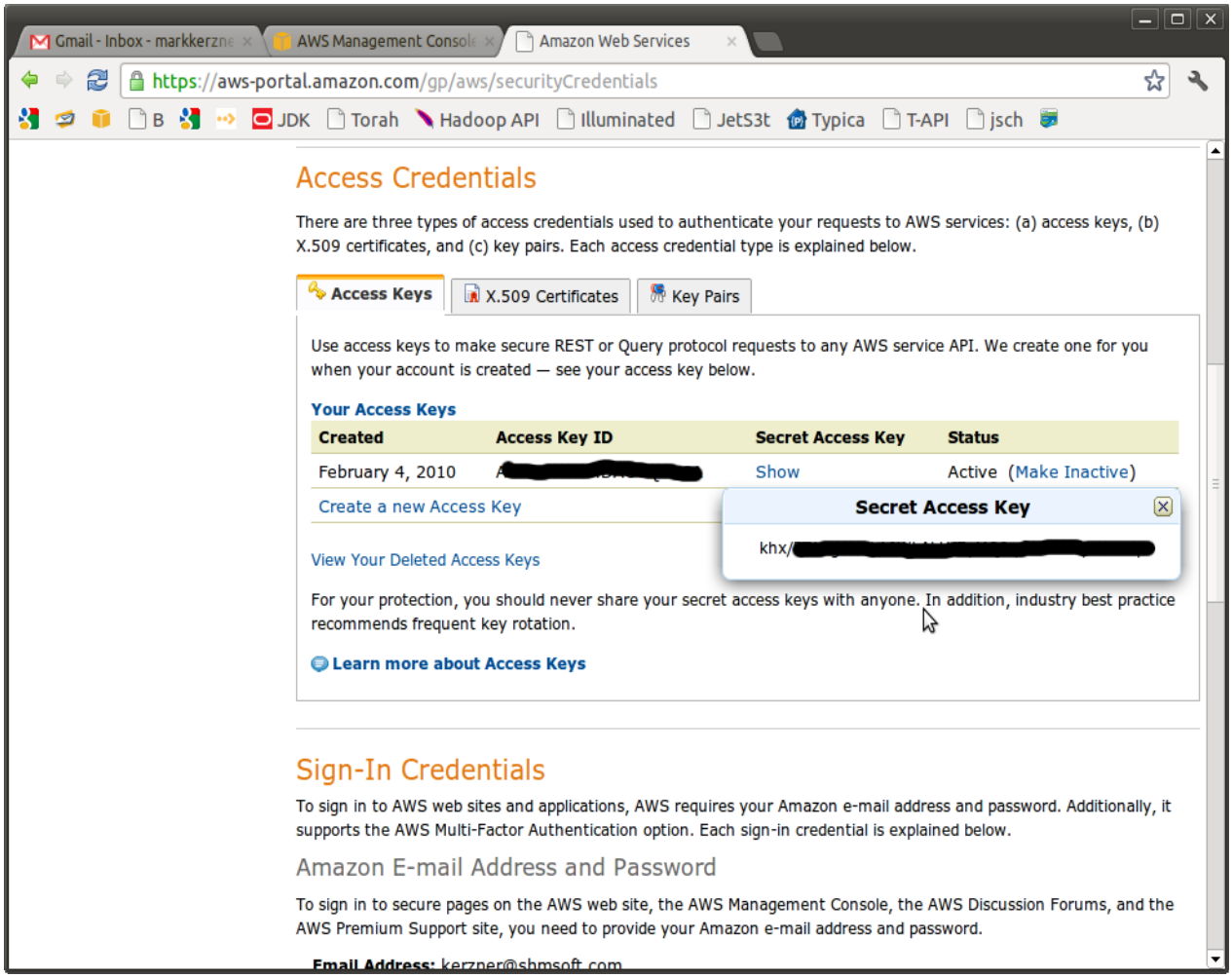

Gmail - Inbox - markkerzne
AWS Management Console
Amazon Web Services
https://aws-portal.amazon.com/gp/aws/securityCredentials
B
JDK
Torah
Hadoop API
Illuminated
JetS3t
Typica
T-API
jsch
Access Credentials
There are three types of access credentials used to authenticate your requests to AWS services: (a) access keys, (b) X.509 certificates, and (c) key pairs. Each access credential type is explained below.
Access Keys
X.509 Certificates
Key Pairs
Use access keys to make secure REST or Query protocol requests to any AWS service API. We create one for you when your account is created — see your access key below.
Your Access Keys
Created
Access Key ID
Secret Access Key
Status
February 4, 2010
Show
Active (Make Inactive)
Create a new Access Key
Secret Access Key
khx/
View Your Deleted Access Keys
For your protection, you should never share your secret access keys with anyone. In addition, industry best practice recommends frequent key rotation.
Learn more about Access Keys
Sign-In Credentials
To sign in to AWS web sites and applications, AWS requires your Amazon e-mail address and password. Additionally, it supports the AWS Multi-Factor Authentication option. Each sign-in credential is explained below.
Amazon E-mail Address and Password
To sign in to secure pages on the AWS web site, the AWS Management Console, the AWS Discussion Forums, and the AWS Premium Support site, you need to provide your Amazon e-mail address and password.
Email Address: kerzner@shmsoft.com

Figure 12.3

12.4 - Now Select AWS and click on the S3 Setup button.

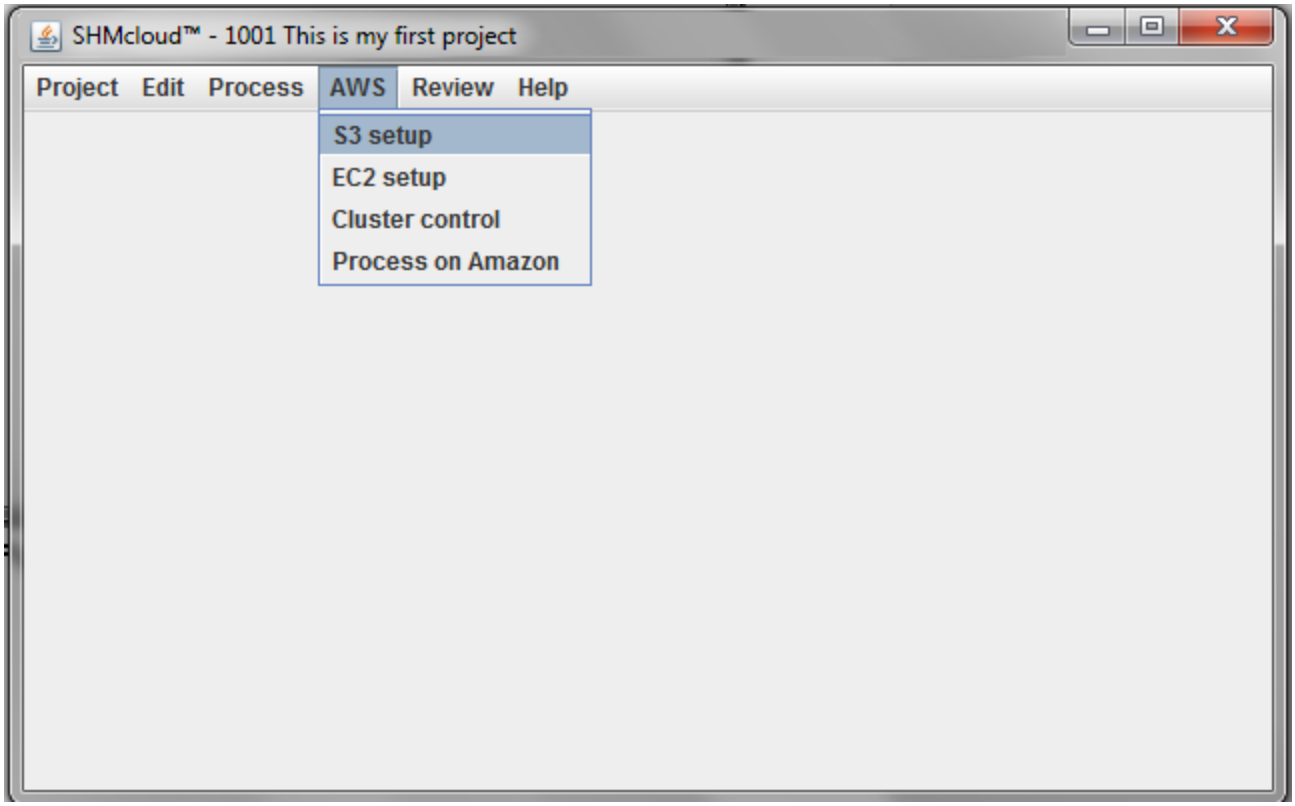


Figure 12.4

12.5 - In Figures 12.2 and 12.3 above we showed you how to find your Access Key ID and your Secret Access Key. Copy and paste your Access Key ID and Secret Access Key respectively into the S3 screen. See below, Figure 12.5.

After you enter your keys, click the “Verify keys” button.

If you do not Verify your keys, then S3 will not work. So you MUST click the “Verify keys” button.

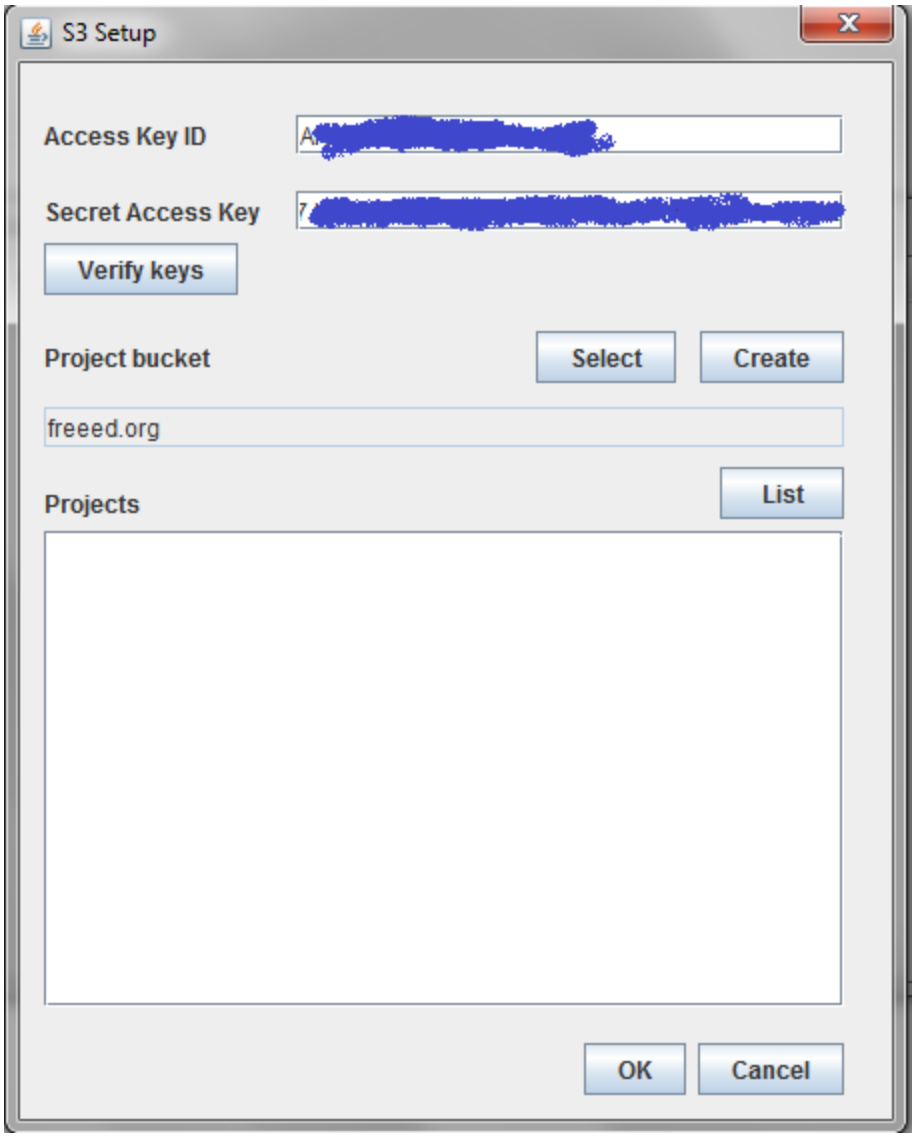


Figure 12.5

12.6 - After you press “Verify keys”, **patiently wait a few seconds**. You should get the following message:

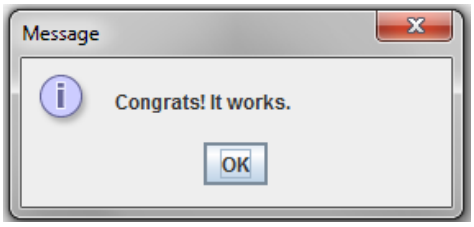


Figure 12.6

Then click “OK” to close the screen.

NOTE: **The process from 12.1 through 12.6 tells Amazon who you are. By entering these keys, you are telling Amazon where to store your data.**

12.7 - We will now create a “**bucket**” or open a pre-existing bucket. Your bucket is like a private folder that belongs to you, only it is located on Amazon. You can use your bucket for anything, not just projects. SHMcloud™ maintains its files there. Within a single bucket you can save an unlimited number of runs for that project, each with a different project name. One bucket can suit all our needs. For example, you may assign a bucket to your department, or to a group working together.

We start by clicking on the “**Select**” button in the S3 screen (Figure 12.5) to choose our project bucket. You will get a list of all your buckets in your Amazon S3 environment. If you have not created any buckets yet, then you will not have any to choose from!

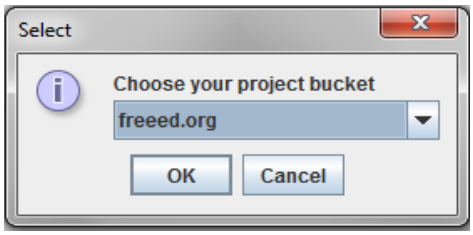


Figure 12.7

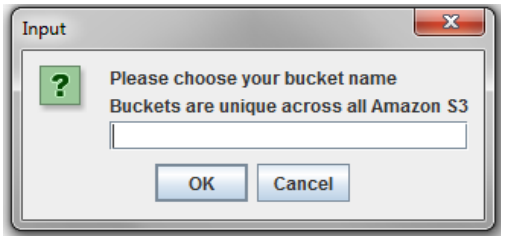


Figure 12.8

You can also create a new bucket for your projects by pushing the “**Create**” button in the S3 screen.

Once you select or create a bucket, it will be shown as your project bucket. Now click on “List” for projects, and you will see a list of your projects stored in this bucket. If the bucket is new, or has not had any projects uploaded to it yet, then the list will remain blank.

If you have projects in your List, then you may choose to select one of them and press OK.

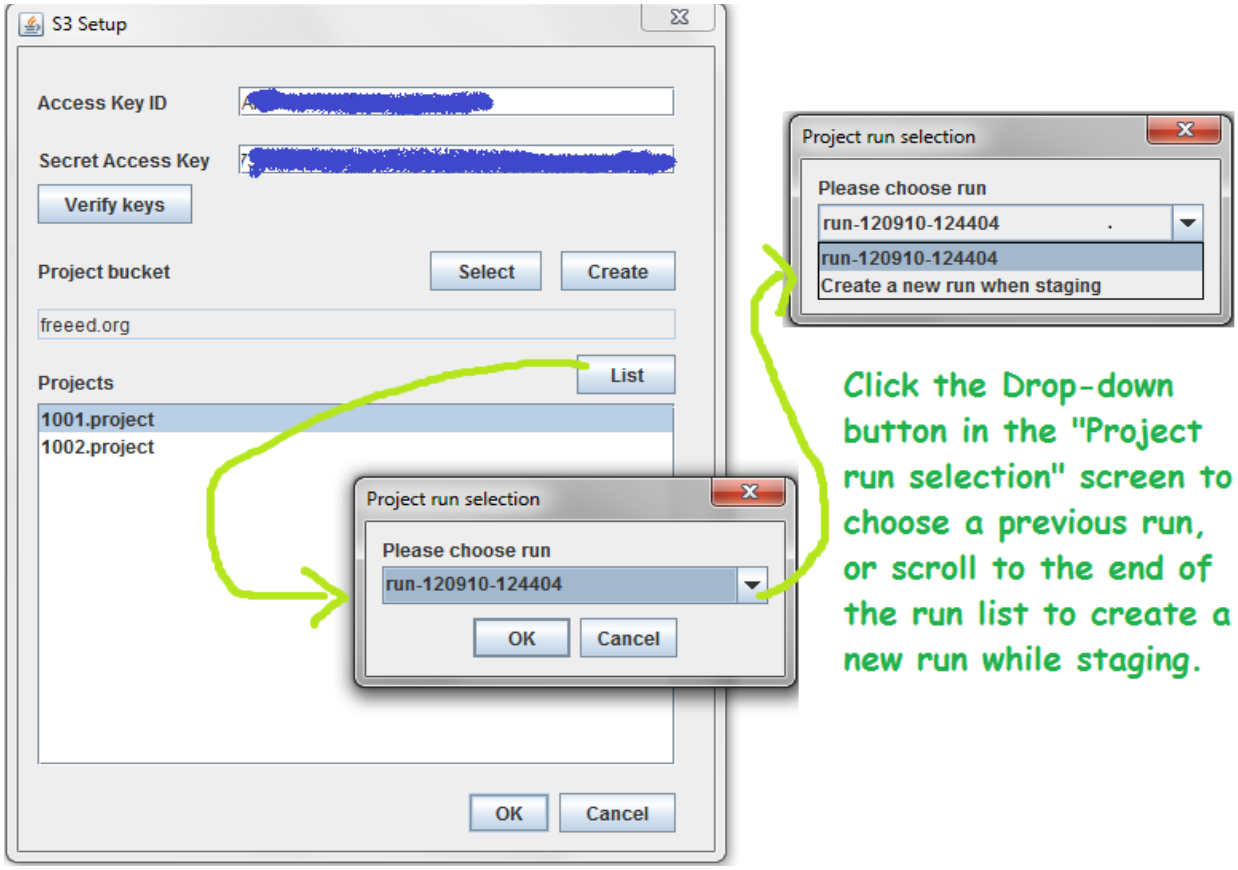


Figure 12.9

When you press OK in the “S3 setup window”, then a “Project run screen” will pop-up if you have already run at least one project.

If you are re-staging your project, then you will want to select the “Create a new run when staging” option that is listed at the end of the “Project run selection” screen. When you run a project, the SHMcloud™ player creates a new subfolder that is based on the "run" timestamp. So **if you use a previously "run" project and don't re-stage, then you will be overwriting that project.** In all other cases you will get new results in a different folder.

Clicking OK in the “Project run selection” screen will open the project and you will see a “Projects Settings” screen similar to **Figure 10.9**, as seen in a previous section of this manual.

Click “OK” to close this screen and move on to the next part of our process.

NOTE: If this is the first time setting things up, there will be no projects listed. No worries, it is just that no projects have been uploaded to your Amazon S3 environment yet. (That will be one of our next steps below.)

Bucket & Project Notes:

⇛ Definition: S3 means Simple Storage Service. So the S3 is your actual storage.

⇛Definition: EC2 stands for Elastic Compute Cloud.

⇛ The project settings are copied from its storage in your bucket onto the local hard drive and the project is then opened. The project opens in the regular way, with the Project Setting dialog coming to the forefront. Fortunately the software takes care of this. We are simply providing this information as an explanation to help our users better understand the process.

⇛ Running your project for the first time will put it into the project List, as seen in Figure 12.9.

⇛ The buckets are unique across Amazon, not unique in your account. Think of it like a URL. In fact, it can be part of a URL, if you make it public.

⇛ Private buckets are invisible, but you can publish buckets or files from within them.

⇛ ***Summary:*** Your **bucket** is like a private folder that belongs to you, only it is located on **Amazon**. You can use your bucket for anything, not just projects. SHMcloud™ maintains its files there. Within a single bucket you can save an unlimited number of runs for that **project**, each with a different project name. One bucket can suit all our needs. For example, you may assign a bucket to your department, or to a group working together. When you run a project, the SHMcloud™ player creates a new subfolder that is based on the "run" timestamp. So if you use the same "run" and don't re-stage, you will be overwriting your data. In all other cases you will get new results in a different folder.

S3 - Abridged steps

(1) Verify Keys
(2) Select or Create Bucket
(3) Click List button then either choose previous run, or scroll to create a new run when staging.
(4) Verify “Project run selection” screen.

Moving Forward:

⇛ Soon we will tell Amazon where to take the processing power from.
⇛ We will learn how to set up a security group on the AWS console.
⇛ We will also discuss the **SHMCloud cluster utilization rules.**

13. Amazon's Strong Security on EC2

We will now learn how to set up access to Amazon's strong security on EC2.

We will be setting up a Security Group, and also Key Pairs. Both security features work independently of each other, which adds to the strength of the security that Amazon offers.

The Key Pairs are called "Pairs" because the user downloads the specific private key, while Amazon keeps the public part of the key.

13.1 - Select AWS Management Console from My Account / Console in the upper right hand corner of your Amazon account.

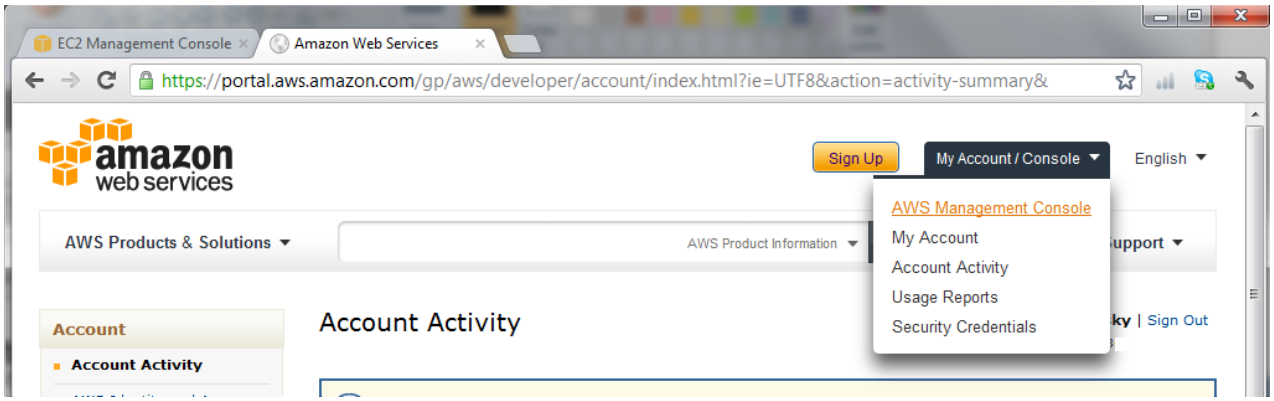


Figure 13.1

13.2 - Then select "EC2 Virtual servers in the Cloud".

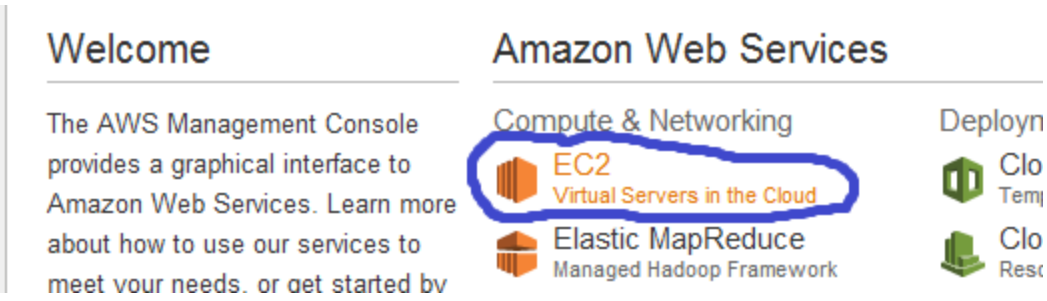

Welcome
The AWS Management Console provides a graphical interface to Amazon Web Services. Learn more about how to use our services to meet your needs, or get started by
Amazon Web Services
Compute & Networking
EC2
Virtual Servers in the Cloud
Elastic MapReduce
Managed Hadoop Framework
Deployn

Figure 13.2

13.3 - Selecting EC2 as seen in the Figure above will take you to a screen that resembles the following image:

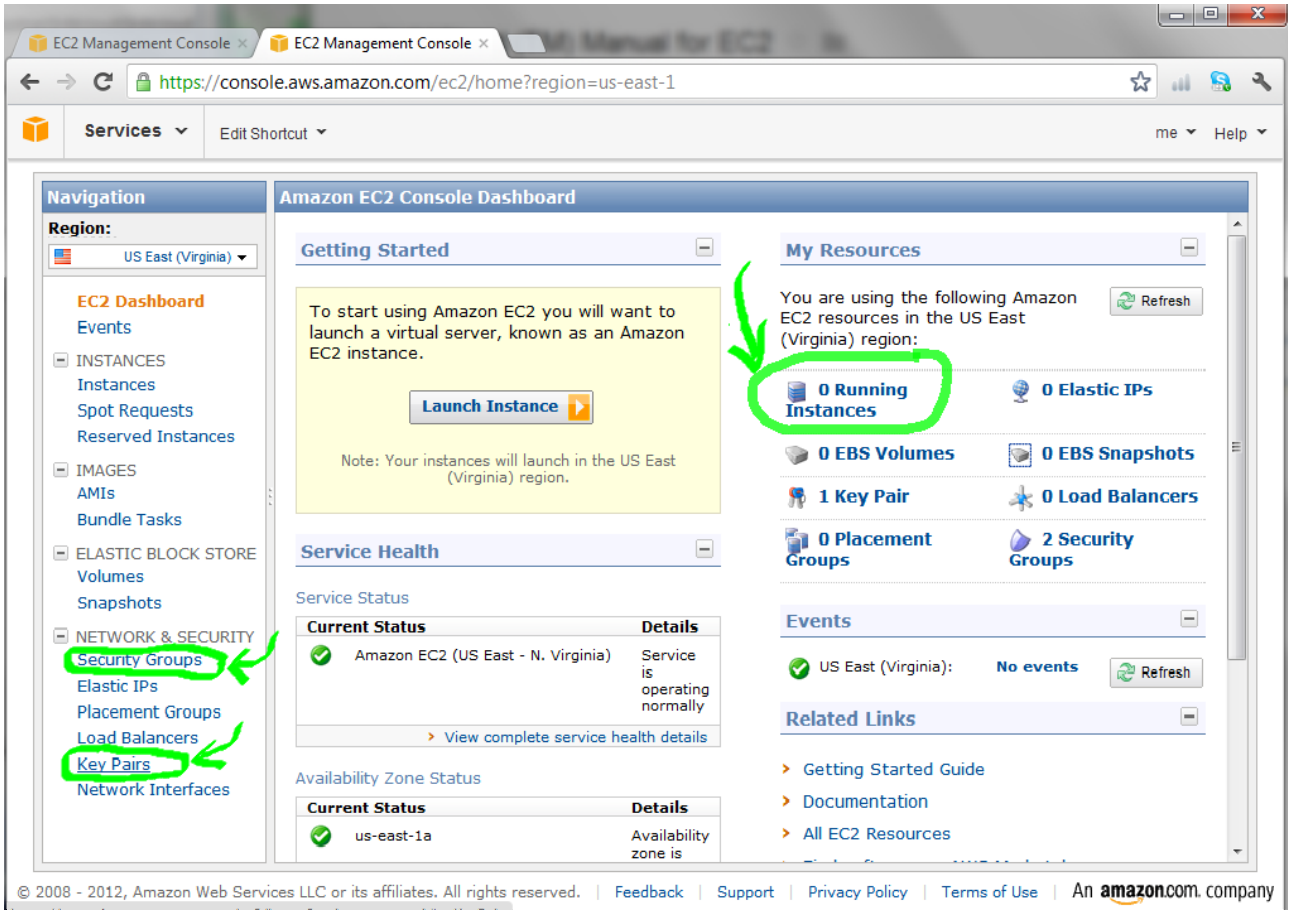


Figure 13.3

As a side note, if you look to the right of the screen under “My Resources”, you will see that currently we have “0 Running Instances”. This is an important observation. It means that we are currently not running any projects. We will discuss this again soon.

At this point we will set up our Security Groups. <u>This is the Firewall</u>. Don’t worry, we should only have to do this one time!

Setting up a Security Group

13.4 - We will start by setting up our Security Group.

⇒ Click on “Security Group”, as seen in Figure 13.3 above. Doing so will open a new screen called “Security Groups”.

⇒ Select “Create Security Group”, as seen below in Figure 13.4. A window (as shown below) will pop up for you to type in the name of your security group, as well as a description for it. I called my security group “hadoop”, with the description “hadoop cluster”. You can call your group by whatever name you choose and give it any description that makes sense for you.

⇒ We will keep the “VPC” selection at the default “No VPC”. You can learn more about other options for setting the VPC by clicking here: http://aws.amazon.com/vpc/.

⇒ Click “Yes, Create”. ***Congratulations!*** You have just created a Security Group on AWS.

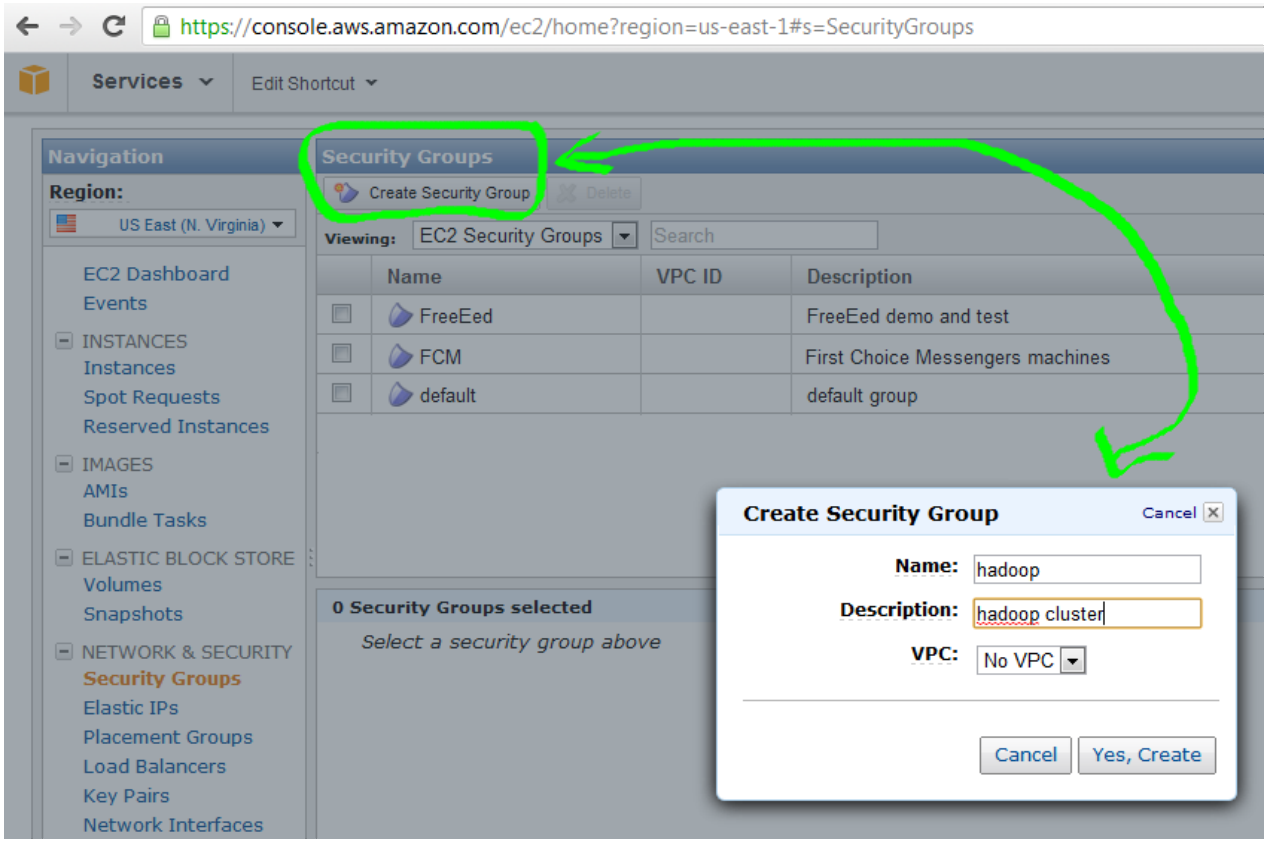


Figure 13.4

13.5 - Now click the button for “Viewing” next to your new security group. A screen similar to Figure 13.5 will open at the bottom of your Security Groups window.

Click on the “Inbound” tab.

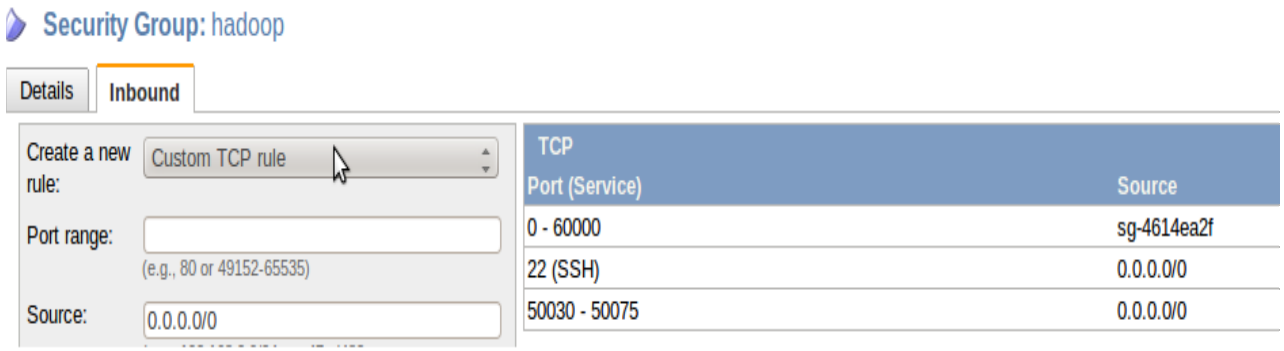


Figure 13.5

You can set permissions for your security group as in the example above, with port 22 open for SSH (remote login) and ports 50030 through 50075 open for Hadoop. If you prefer, you can set it for more restricted access, for example, you can limit access to your computer’s IP only.

Setting up Key Pairs

13.6 - Now we will select “Key Pairs” as seen in the lower left side of the above Figure 13.3 in the Navigation bar.

“Key Pairs” are one of the many security features that SHMcloud™ provides for our users in order to guarantee the protection of sensitive data.

If you do not already have any Key Pairs, then your next screen will show no keys.

⇒ Click on the “Create Key Pair” tab. A screen similar to Figure 13.6 will open. You may call the Key Pair by whatever name you want. I have chosen to call my Key Pair “shmcloud”.

Please note that the Key Pair Name is case sensitive, and even a blank space at the end will count as a character.

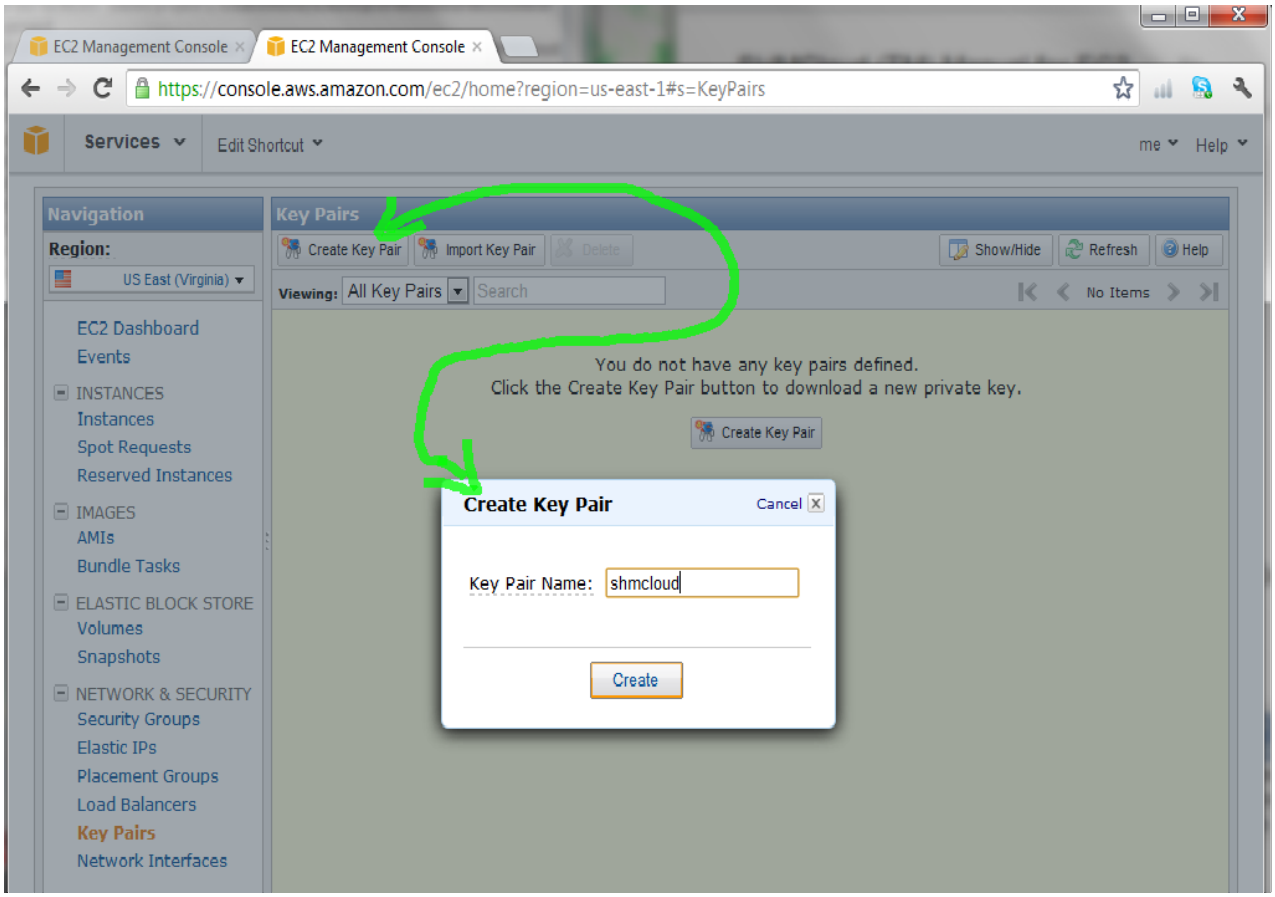


Figure 13.6

13.7 - A screen will pop up telling you that you have created a key pair by the name which you have given it in the previous step.

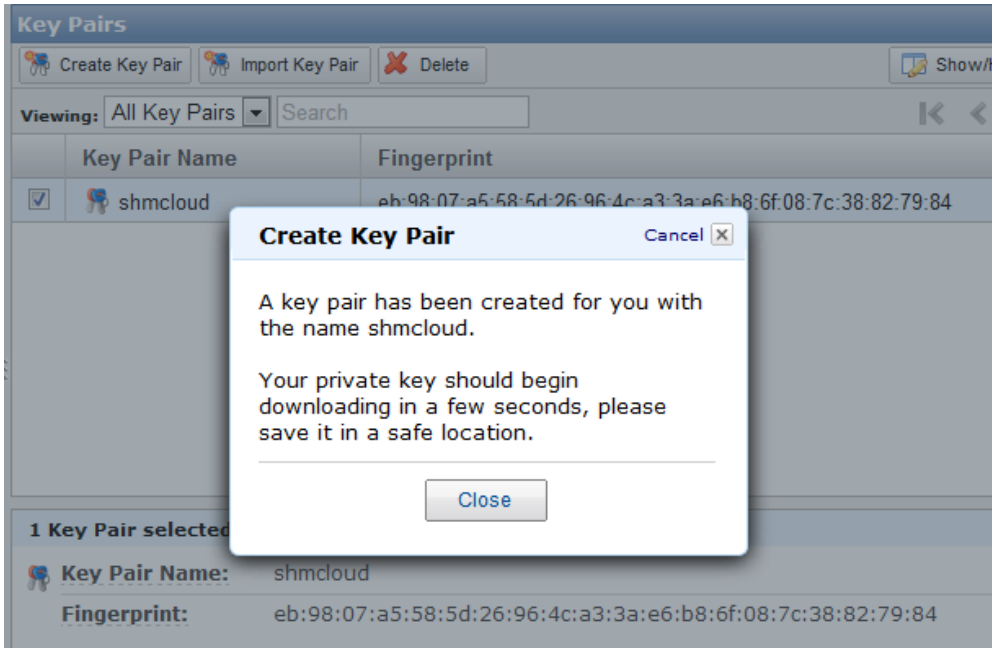


Figure 13.7

⇒ A PEM file will download onto your computer. Allow it to download. The PEM file will contain your private key. As we explained earlier, Amazon keeps the other half of your key pair - the public identifier and name. The public identifier that Amazon holds and the information contained in your PEM file work together like a super lock, which helps to enhance the security of your projects.

⇒ Open the PEM file.

⇒ Select and copy ***(CTRL+C)*** the entire contents of the file, **including all of the dashes** before and after the Begin and End lines.

⇒ IMPORTANT: For maximum security the PEM file can only be download one time, on creation of the key pair, as seen above in Figure 13.7. If your system gets reset you will not be able to access that same key again, unless you saved the file (in a secure location, of course).

Otherwise you will have to set up a new key pair by repeating steps 13.6 to 13.7.

Preparing your EC2 *(Elastic Compute Cloud)* for processing

13.8 - Now go back to your SHMcloud™ Player. Select **AWS**, and then click on **EC2 Setup**.

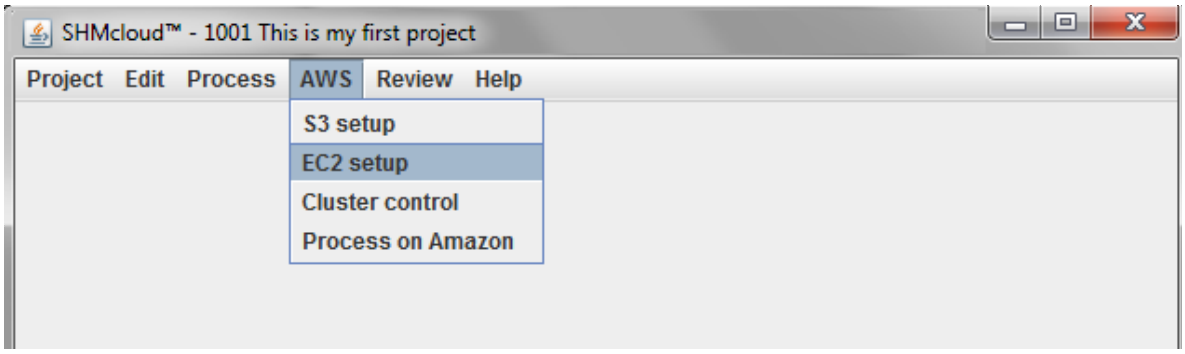


Figure 13.8

13.9 - The screen below will pop up.

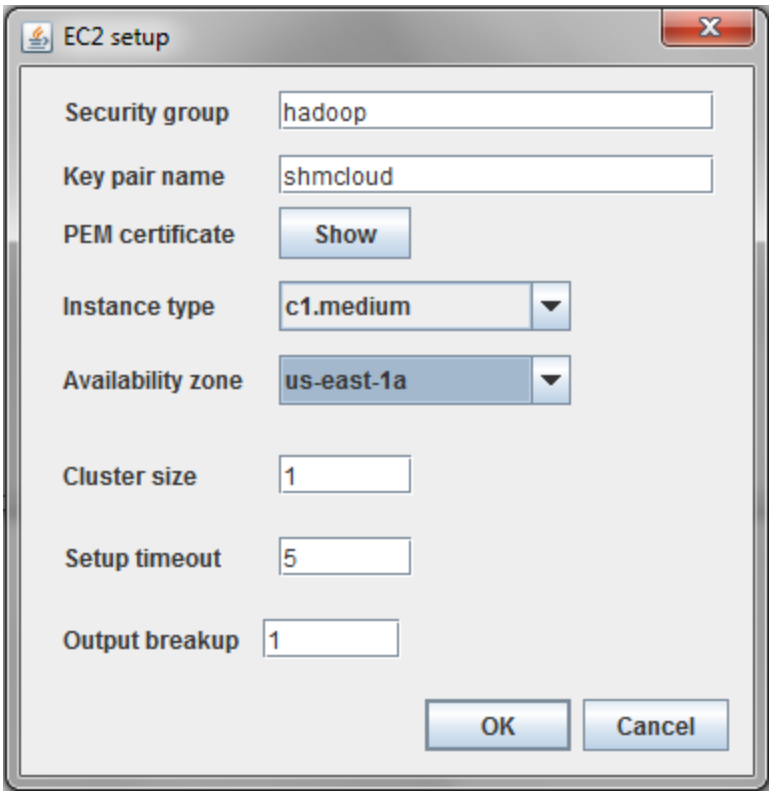


Figure 13.9

In Figure 13.4 we called our **Security group** "hadoop". Then in 13.6 we gave our **Key pair** the

name “shmcloud”. Now, as seen above in Figure 13.9 we enter those names accordingly into the EC2 setup screen.

⇒ You must type the Key pair name and the Security group exactly the way you named them on creation. **If you make a typo or even put in an extra space in either of those entries here, then you may not be able to run your project.**

⇒ In the EC2 setup screen, click the **“Show”** tab that appears next to **“PEM certificate”** ONE TIME, and wait about 45 seconds.
⇒ A blank screen called “PEM Certificate” will pop up.
⇒ Click your mouse into the empty space in that window and then paste ***(CTRL+V)***, which should paste the information that we just copied from our downloaded file, as explained above in Section 13.7. See below, Figure 13.10. For security reasons I blanked out most of my key so that the reader cannot copy my private PEM key!

If your PEM key did not paste when you clicked *CTRL+V*, then please repeat the steps in Section 13.7 above, and retry.

⇒ Once copied, Clicking “OK” in the PEM Certificate screen will save your setup and close the PEM Certificate window.

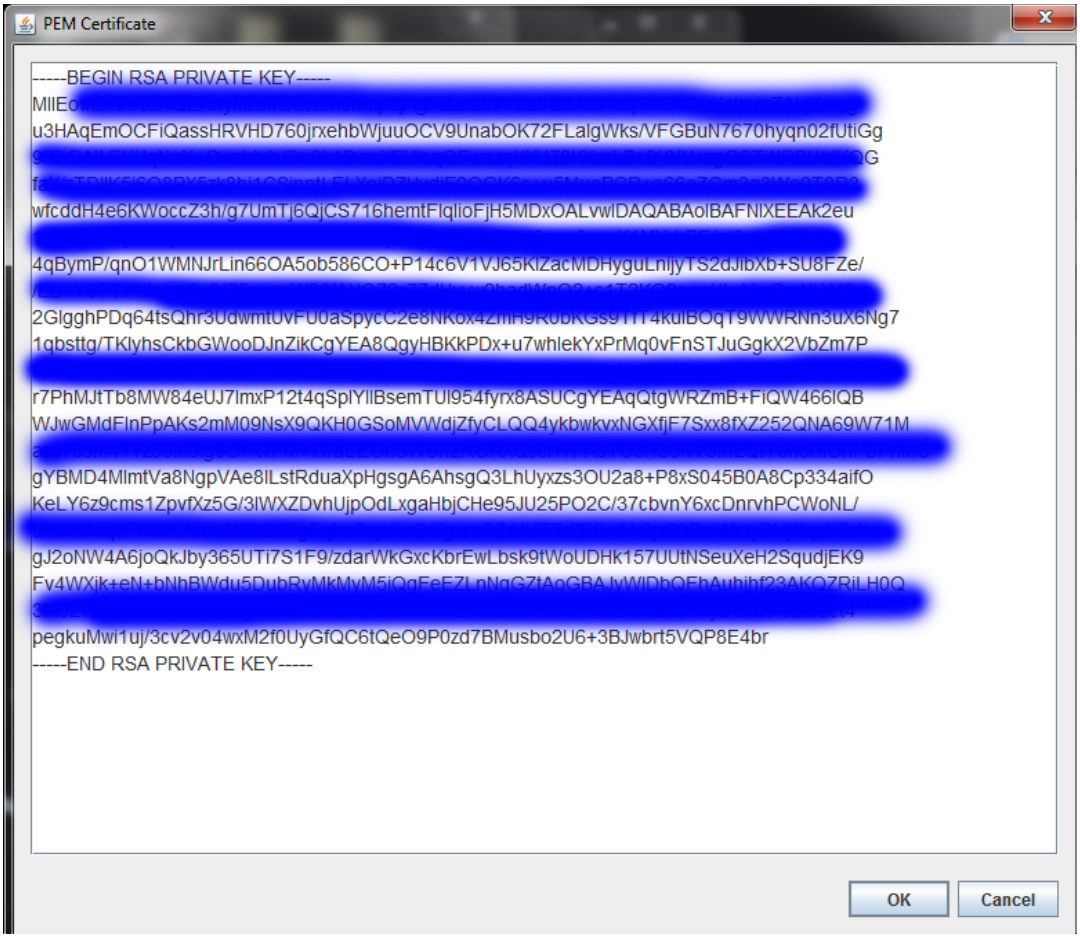


Figure 13.10

NOTE: Setting up your **Key Pairs** and your **PEM Certificate** are security measures that need only be done once. You should not have to redo those steps unless you delete SHMcloud™ from your computer or wish to start over on a different computer.

Let's discuss some of the other options that we can find in the EC2 Setup screen, as seen in Figure 13.9, above.

⇒ The **Instance type** will either be medium (as seen above), or large. The "instance" refers to the computer size on Amazon. There is no option for a small instance because if you have a small project then you will be running your project **Locally** on your computer. Afterall, if it is a small project then why waste money by running it in the cloud?

⇒ The tab that shows the **Availability zone** offers several option for where your project will physically be running. These zones are where the actual Amazon computers are located. Just choose one randomly. If Amazon is running too many projects at the same time in that location and no machines are available, then you will get a message telling you to try a different zone.

This doesn't happen very often, but now you know how to control things if this does happen.

⇒ The **Cluster size** tells Amazon how many of their supercomputers you would like to use to process your project. The more computers you use, the faster your project will complete. However there is an added fee for each computer that you include per run. Depending on the size of your input data, you should carefully decide if the added speed will outweigh the cost of the extra computers, before you determine what the best choice is for your job.

Guidelines for cluster size

1 instance
One can run a complete cluster on a single EC2 instance for testing, selecting cluster size as 1. In that case, all Hadoop services run on that one instance.

2-10 instances
One instance (the first one) is used as a "master." It controls the HDFS file system and the organizes the work of the other instances. All other instances are used as workers, or "slaves". The store the HDFS file data and perform actual eDiscovery work.
The 5-10 nodes is the recommended configuration during the inital testing period.

11-50 instances
One instance is used as an HDFS file system controller (called namenode), another one organizes and controls processing jobs (call jobtracker), the rest are workers (slaves).

⇒ **Setup timeout** allows the user control over how much time to give the cluster to begin. If the cluster does not start in that amount of time, then there may be a problem with the EC2 setup. Five minutes is a safe amount of time to set for the cluster to begin.

⇒ **Output breakup** allows the user control over how many zip files the output should be divided into, for convenience of handling.

This completes our setup of the EC2 screen for processing on Amazon using SHMcloud™ .

You may now click OK to exit the EC2 setup screen.

14. Cluster Control - How to Turn on your Cloud Computer & Run Your Project on Amazon

14.1 - We will now we open the cluster control screen, Figure 14.1.

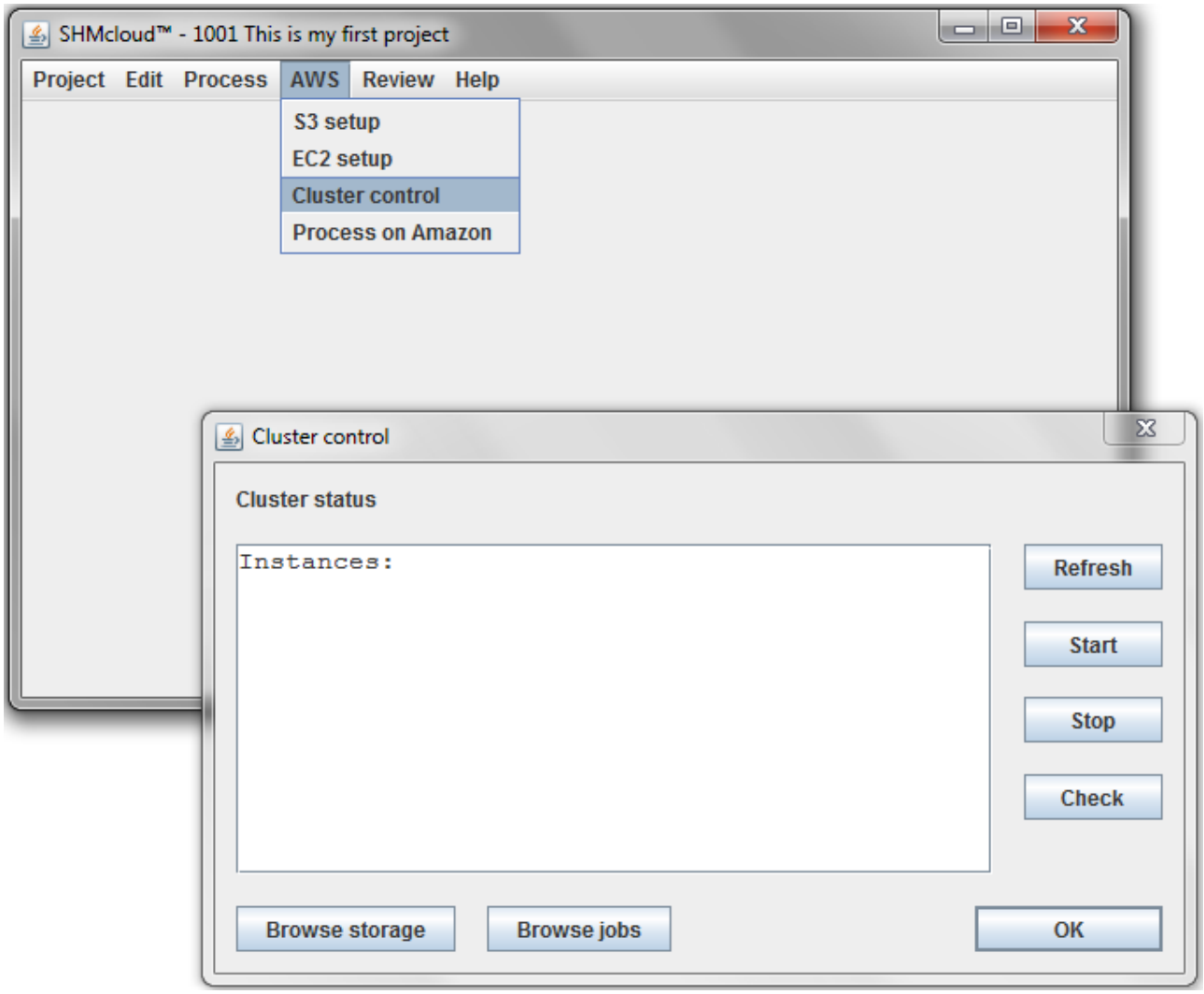
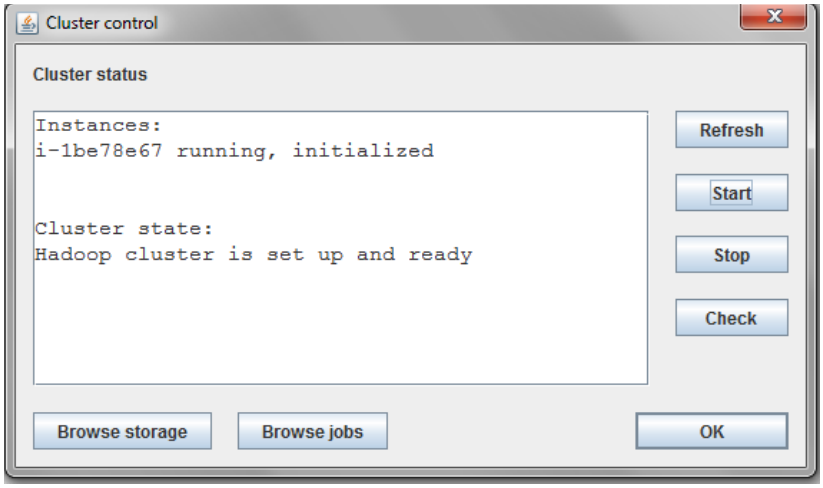


Figure 14.1



...

The Cluster control starts the cluster on Amazon. Think of the cluster as being a super computer. In essence, the cluster is really a bunch of computers set up together to do the tasks that you assign it.

Clicking on Start, Figure 14.1, will be like turning on your very large computer. But this computer happens to exist in a cloud run by Amazon!

⇛ Dont rush. Click on the buttons one time, **and wait for it.** Clicking Start more than once might turn on more than one Cluster instance. So just be patient and wait while things turn on.
⇛ Click Start in the Cluster control screen. It should take about 5 minutes to begin.
⇛ A message will come up telling you that the your Cluster has begun.
⇛ Click OK to exit the message, and then click OK again to exit the Cluster control screen.

There is a lot of functionality happening in Figure 14.1

⇛ ***Refresh*** - refreshes the status of the cluster

⇛ ***Start*** - starts the cluster. This includes starting the EC2 instances; once the instances start and accept connection, putting the required SHMcloud software on each instances, setting up the Hadoop cluster, starting Hadoop services, and running a sample job to verify the operation.

⇛ ***Stop*** - stops the clusters and disposes of the cluster machines.

⇛ ***Check*** - run the cluster verification by running a sample job.

⇛ ***Browse storage*** - opens a browser to the files system (HDFS) on the cluster

⇛ ***Browse jobs*** - opens a browser to Hadoop jobs: scheduled, running, and completed.

14.2 - Everything is all set up. It is now time to process your job on Amazon's super computer.

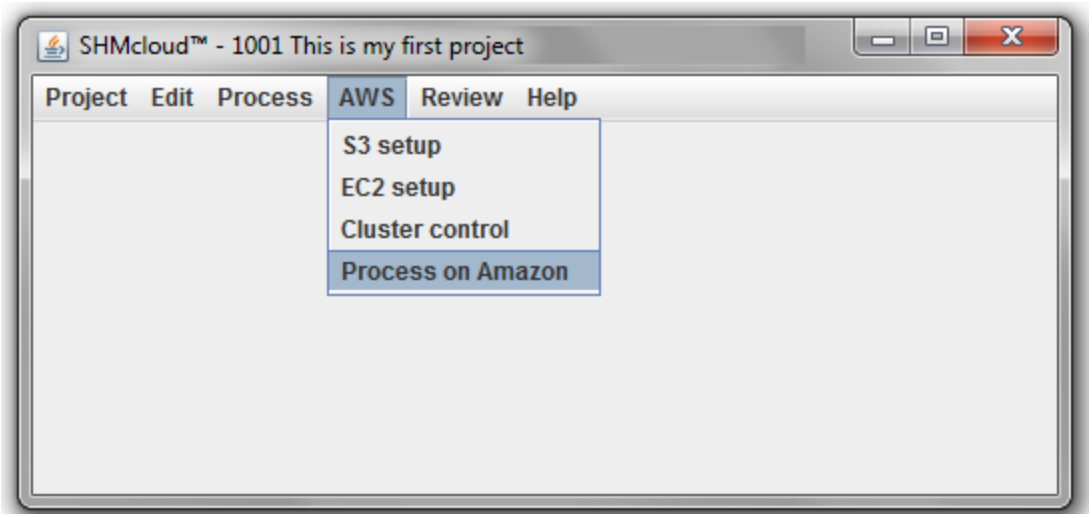


Figure 14.2

Select "Process on Amazon" from the AWS selection in the SHMcloud™ menu.

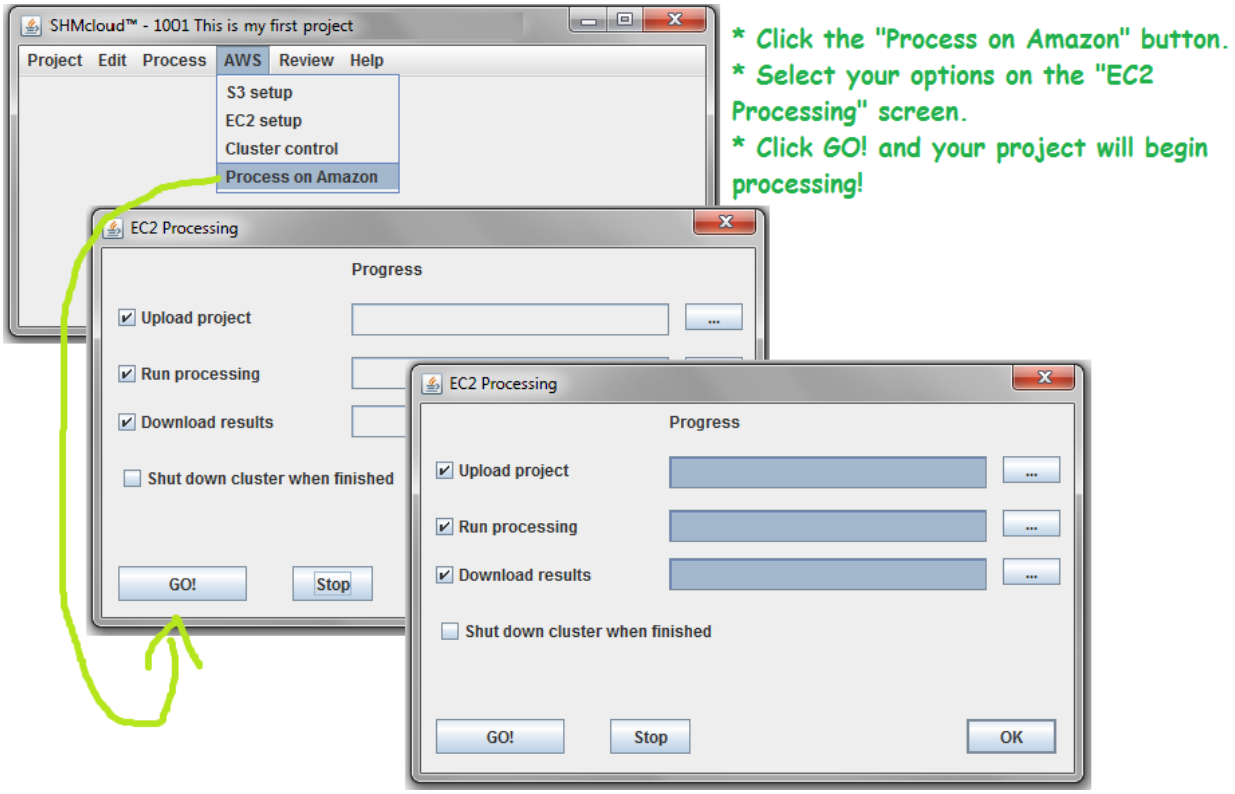


Figure 14.3

⇛ Select your options from the “**EC2 Processing**” screen.

⇛ Click GO, and your project will begin processing.

14.3 <u>Other Notes on this screen:</u>

⇛ Notice the “...” next to the processing lines in the EC2 Processing screen above in Figure 14.3. If you click on it, then details of that particular step will be revealed.

⇛ The **Stop** button in the **EC2 Processing** screen will stop your job from processing. HOWEVER, the cluster will still be on, and Amazon will continue to charge for the time. You may turn off the cluster by pushing **Stop** in your **Cluster Control** screen, as seen in Figure 14.1

⇛ You may keep an eye on the progress of your job by keeping the EC2 Processing screen open for the duration of the run. However, even if you close the screen, your job will continue to process until it ends on its own, or gets terminated due to some unknown reason.

14.4 Shutting Down the Cluster

IMPORTANT: Your Amazon account is charged by the hour for running time, so don't forget to stop the cluster once you are done.

⇛ When your job finishes processing, the Amazon cluster will continue to run. There is no automatic shutoff on Amazon AWS at this time. Shut down the cluster by clicking on the "Stop" button as shown in Figure 14.1. This will shut off the Amazon computers, and you will stop being charged.

How can you determine that the cluster really turned off?

Earlier, as seen in Figure 13.3, we we showed you **"0 Running Instances"** in the upper right corner of the **EC2 Management Console**, and pointed out that this was something very important to notice.

Go back to your **EC2 Management Console**. If the number next to **"Running Instances"** is anything greater than a zero, then you are still running the cluster, and Amazon will be billing you for the time. You can force a shutdown of the cluster from within the EC2 Management Console, however it is best to follow the SHMcloud ™ guidelines by shutting down your cluster in the Cluster Control by pressing **"Stop"**, shown in Figure 14.1. Following proper shutdown guidelines will help to maintain the integrity of your output.

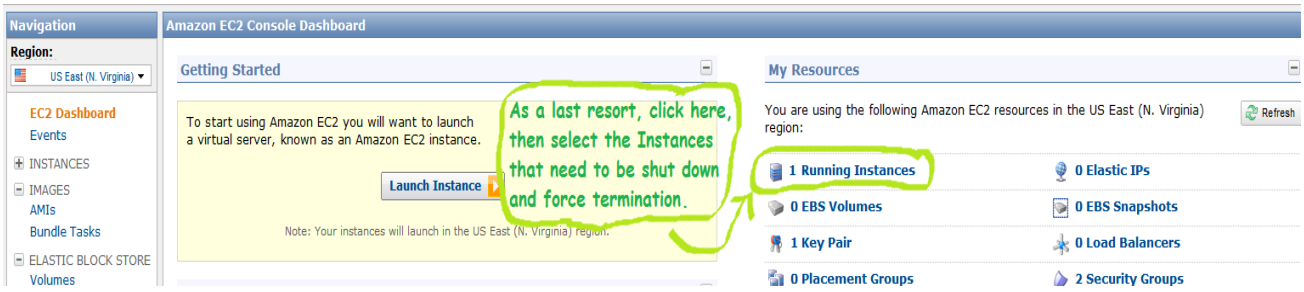

Navigation
Region:
US East (N. Virginia)
EC2 Dashboard
Events
INSTANCES
IMAGES
AMIs
Bundle Tasks
ELASTIC BLOCK STORE
Volumes
Amazon EC2 Console Dashboard
Getting Started
To start using Amazon EC2 you will want to launch a virtual server, known as an Amazon EC2 instance.
Launch Instance
Note: Your instances will launch in the US East (N. Virginia) region.
As a last resort, click here, then select the Instances that need to be shut down and force termination.
My Resources
You are using the following Amazon EC2 resources in the US East (N. Virginia) region:
Refresh
1 Running Instances
0 Elastic IPs
0 EBS Volumes
0 EBS Snapshots
1 Key Pair
0 Load Balancers
0 Placement Groups
2 Security Groups

Figure 14.4

To **force a shutdown** of your Instances on the cluster, click on "Running Instances" in the EC2 Management Console, see above Figure 14.4. You will see exactly what instances are running. You can then select them and terminate. This will be a proactive forced shutdown of the cluster so that Amazon will no longer bill you for the time usage. However, as stated in the previous paragraph, it is best to shut down the cluster from within the SHMcloud™ software. The forced shutdown option should only be invoked if you are having difficulty shutting down the cluster by following the proper steps.

14.5 - Reviewing your output after running your project on Amazon

Reviewing output after running your project using Amazon's supercomputers is done by following the same steps as we have taken to review output from a local run, as seen in section 9, and sections 10.19 through 10.21.

There are, however, a few differences in the output itself.

Click "**Review**" in the SHMcloud™ menu, followed by "**Open output folder**".

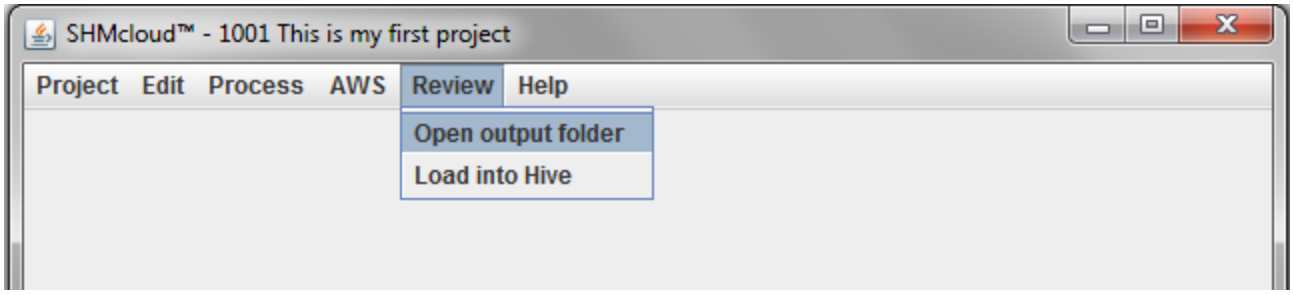


Figure 14.5

14.6 Earlier, as we discussed at the end of section 10, opening the output folder after a local run rendered three files or folders called "metadata", "native" and "report".

Now, as seen below in Figure 14.6, there is a file called "load-00000", and several zip folders starting with the name "attempt_....".

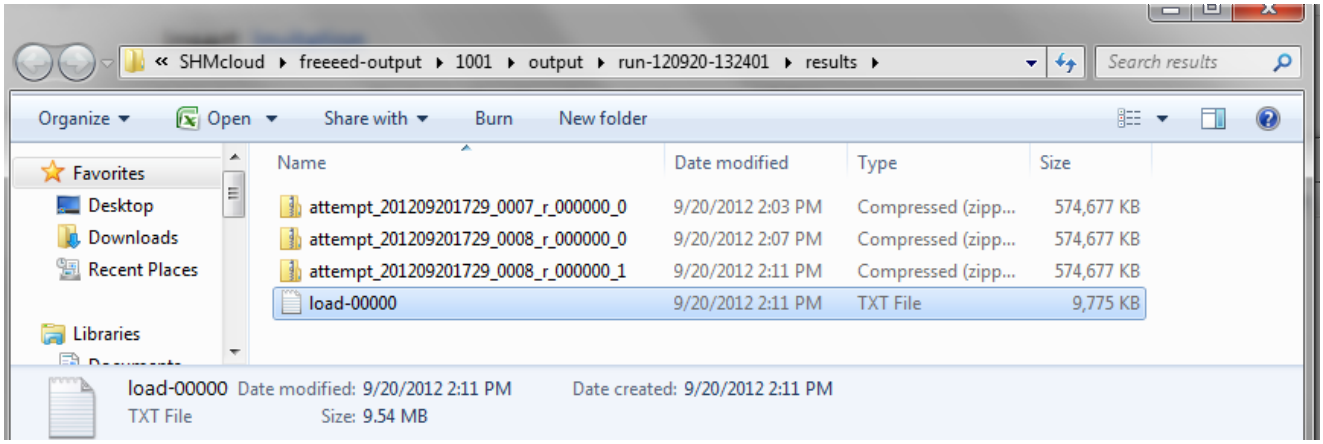


Figure 14.6

The file called "load-00000" is equivalent to the "metadata" file that we reviewed earlier, and should be opened and read using the same methods that we discussed in section 9 of this manual.

As we discussed in the “Notes & Warnings” section at the end of section 10, the “native” folder that is produced by a **local run** is a zipped folder. It contains all extracted native files, including emails and text extracted from them, as well as “exception” files that could not get processed for any reason. Essentially it is everything that this project processed.

Similarly, the zipped folder(s) produced by an **AWS run** called **“attempt_.....”** contains all extracted native files, including emails and text extracted from them, as well as “exception” files that could not get processed for any reason. Essentially it is everything that this project processed.

As seen in our example above, there can be many “attempt_...” zip folders produced by a single project. The number of output folders can be determined by the user in the EC2 setup screen, seen below in Figure 14.7. The user need simply enter a number for “Output breakup”, to decide how many output folders should be created during the run of their project. If the user does not enter a number that is greater than 1, then the SHMcloud™ software will break the output into as many folders as it needs to during processing.

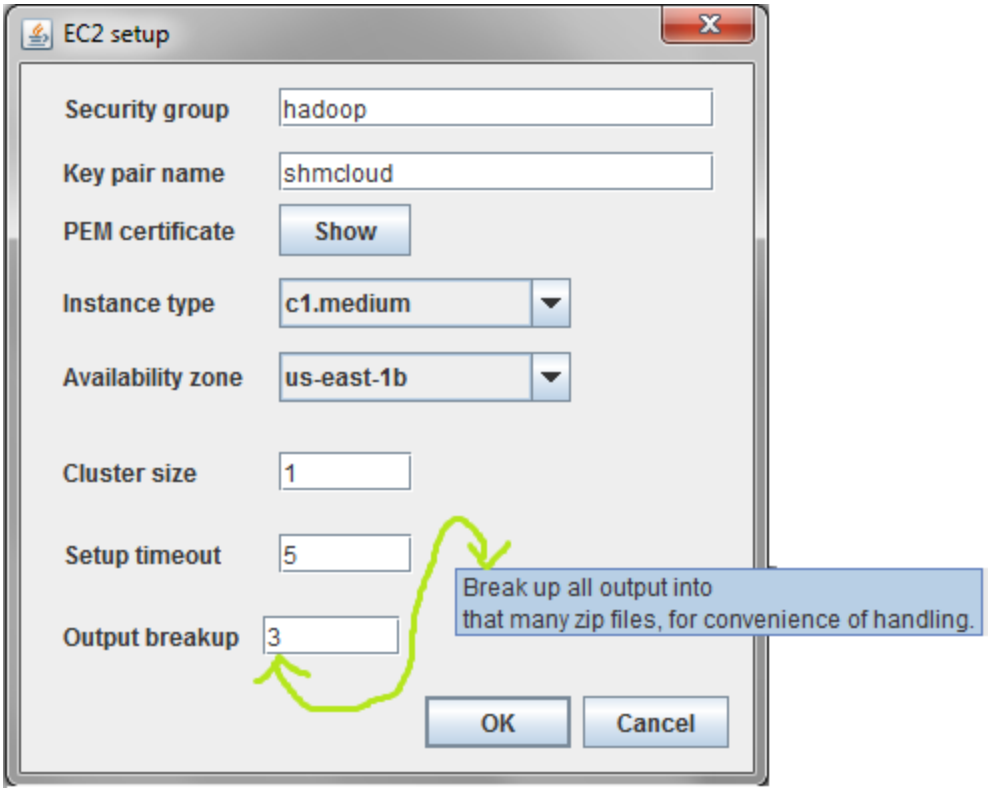


Figure 14.7

15. Creating Projects With Specialized Searches

Now that we are comfortable with creating and processing our projects both locally and in the cloud, it is time to get down to the real business of searching through our output for specific text strings, dates, etc.

SHMcloud™ creates data that can be used with the Apache Solr search server. We will follow through the basic steps for creating usable data and include guidelines on how to use Solr in conjunction with your SHMcloud™ Player.

For more information about Apache Solr, please link here: http://lucene.apache.org/solr/

Installing Solr on your computer for use with SHMcloud™

Before you can use Solr for searching, it is necessary to download and install it onto your computer. For your convenience we have included the following simple steps. ***It is necessary to include steps 1-7 for the initial setup of Solr on your machine.***

1.) **Download** the solr installation package, version: apache-solr-3.6.1.
The url for the direct download is: http://apache.online.bg/lucene/solr/3.6.1/apache-solr-3.6.1.zip

2.) **Unzip** the file.

Steps 1 and 2 need only be done once, unless you are updating to a different version of apache-solr, or changing machines.

3.) From within your SHMcloud™ directory on your hardrive, go to the **Config folder**.

Copy the **config/schema.xml** configuration file to **apache-solr-3.6.1/example/solr/conf** , which you just unzipped in step #2. Select the **copy & replace** option, if necessary.

Step 3 should also only need to be done once, even if you upgrade to a later version of SHMcloud™, unless there is an instruction to repeat this step from within SHMcloud.

4.) Go to **apache-solr-3.6.1/example** on your harddrive.

5.) Double-click "**start**" to start Solr: java -jar start.jar

6.) Check the output for errors. (If you have a CMD screen opened, any errors should appear there.)

7.) Go to: http://localhost:8983/solr/admin

Steps 4-7 will need to be repeated every time that you restart your machine.

Notes:

⇛ http://localhost:8983/solr/admin **is local to your personal machine. You can use Solr Search for searching specifics of your output there. Anytime you restart your computer you will have to turn Solr back on. If your computer is always on, then Solr will remain on.**

⇛ **The SHMcloud™ player does NOT automatically turn Solr on for you. Solr needs to be opened prior to your run in order for your output text to be written into it.**

⇛ **Once installed, turning Solr on is simple. As shown in the steps above, find apache-solr-3.6.1/example on your hard drive and double-click "start" to start Solr. There will be no bells or whistles, it will simply just turn on.**

⇛ **The output from your run will remain in Solr until you process your next job with Sol turned on. It will then write over the previous project's output.**

How to run projects in conjunction with the Solr Search Server:

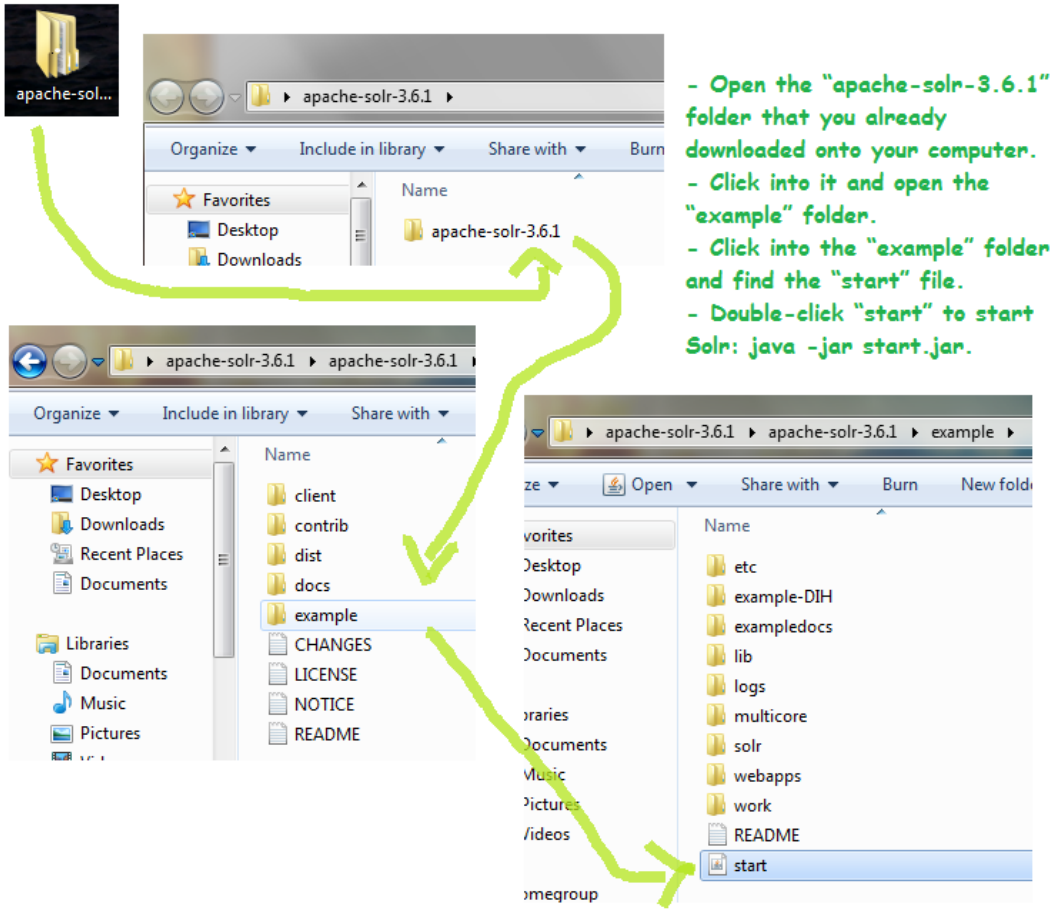
➡ Please note that the current release of SHMcloud™ does not have Solr installed for Cloud processing. However, the SHMcloud™ Player has been designed to include Solr as a search tool for local processing.

As outlined in the previous steps, you should have already installed apache-solr-3.6.1 onto your hard-drive. **But remember, Solr will not work if you do not turn on the Solr machine prior to your project run.**

➡ If you have not yet installed apache-solr-3.6.1 onto your computer, then please follow the steps outlined in the previous section.

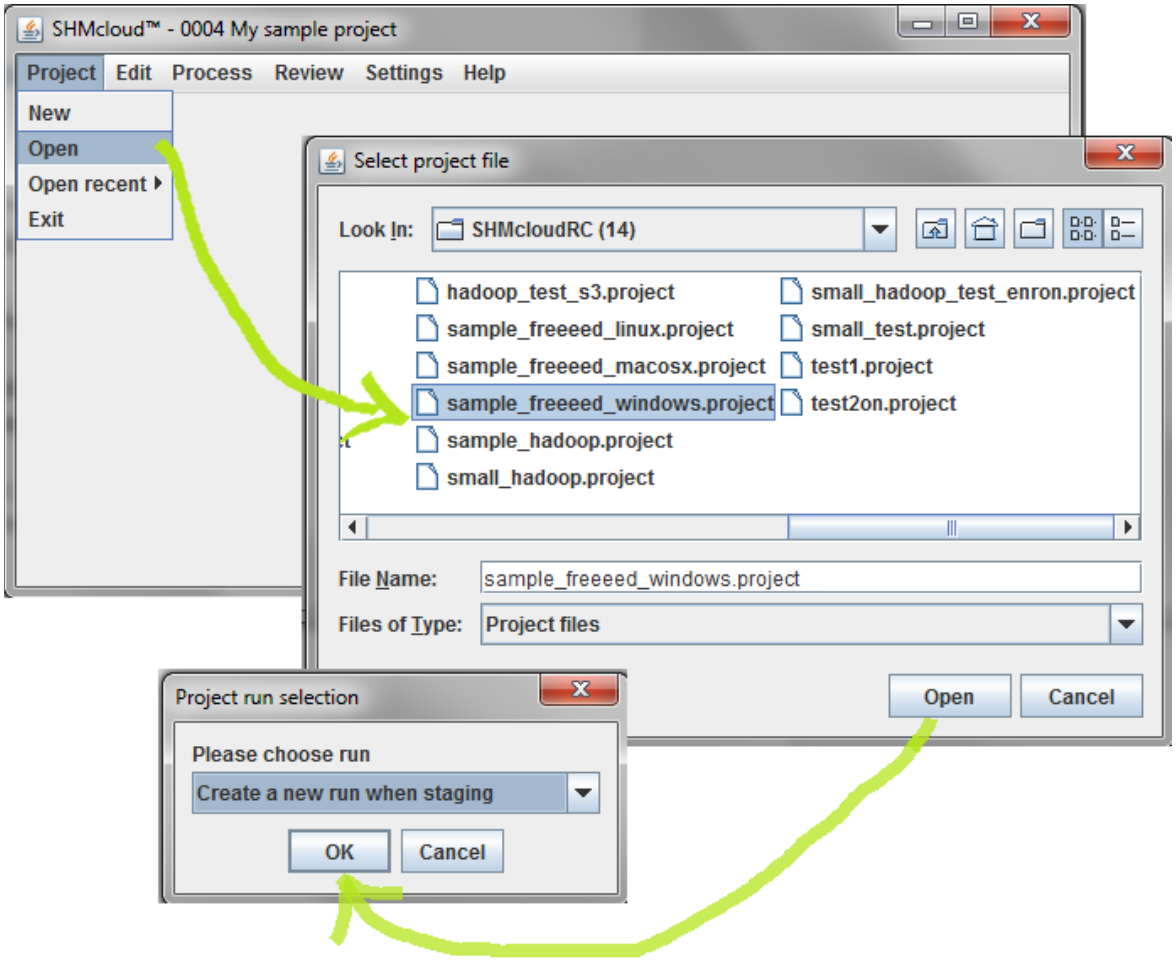
Turning on your Solr Search Server

➡ If you have have not yet turned Solr, or if your machine has been restarted, it is necessary to turn Solr on prior to processing. You may follow the steps in the previous section for turning Solr on, or you can simply follow the steps outlined in the diagram below.



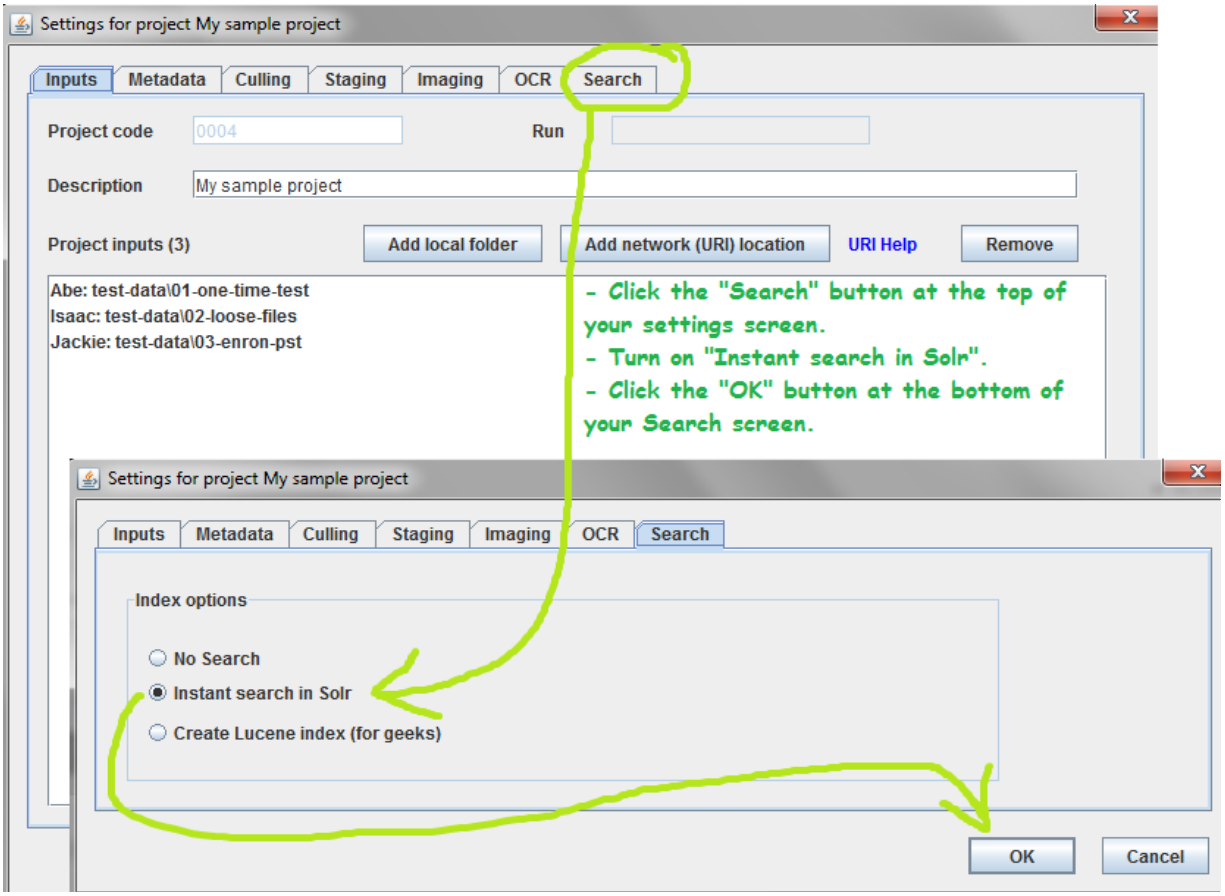
➡ Prepare your SHMcloud™ project run as outlined in the earlier sections of this manual.

For the sake of convenience, the following process is best explained by using the test project that we first used when we tested out our SHMcloud™ player, earlier in this manual.



➡ Open the sample_freeeed_windows.project, and select OK to create a new run while staging. If you decide to choose a previously run project, then the output in that folder will be overwritten.

As seen earlier, your Settings screen will pop-up after you select the Run option for your project.



➡ Click the “Search” button at the top of your settings screen.

➡ Turn on “Instant search in Solr”. The default Search option is “No Search”.

➡ Click the “OK” button at the bottom of your Search screen.

Note:

➡ Currently the option to “Create Lucene Index (for geeks)” is not meant to be used by our regular users.

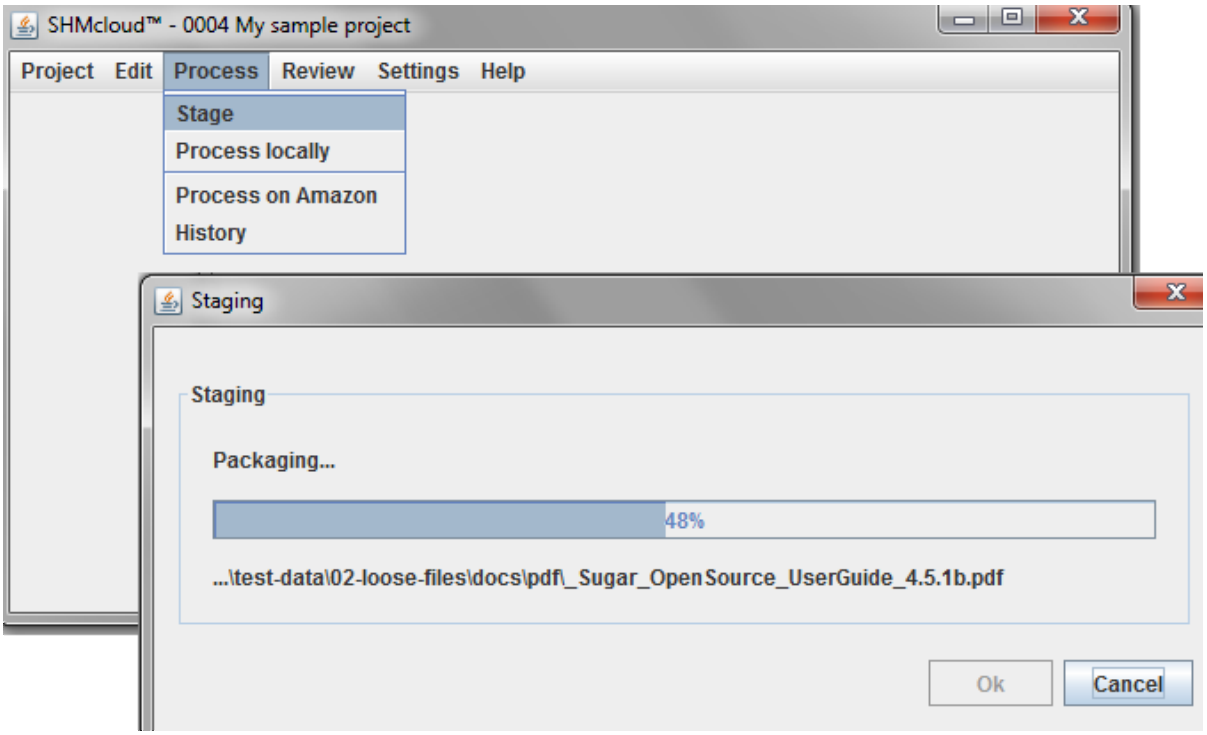
➡ If you choose “Create Lucene Index (for geeks)”, then you will need another program that can browse through it. The file will be created within your SHMcloud™ directory at SHMcloud/lucene_index.

➡ *Feature in the works:* Our plan is to design this option into a usable feature for our regular users to make it easier to move data in and out of Solr, without having to reprocess the job.

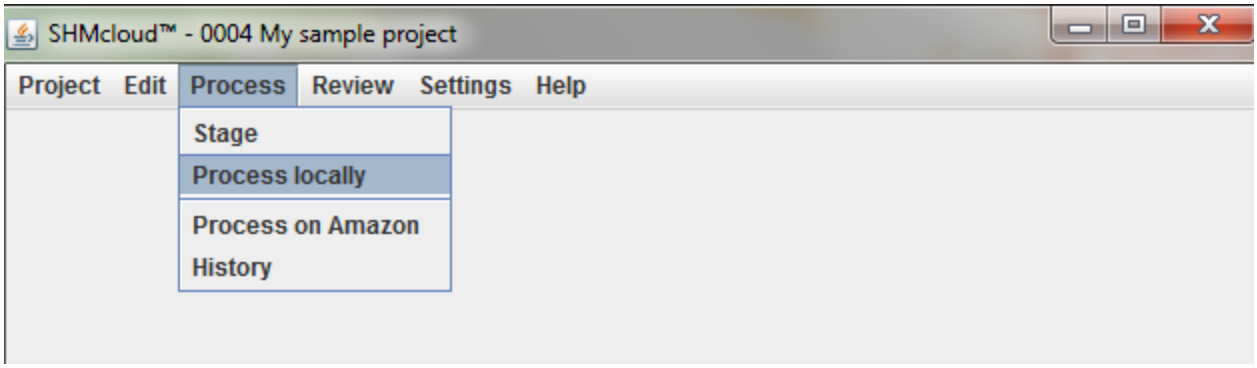
➡ For now let’s stick with the second option and select “Instant search in Solr”. When this option is selected, the documents are sent directly to Sorl. The url listed inside the Solr option should be: http://localhost:8983.

Now we are set to process our project the way that we normally would.

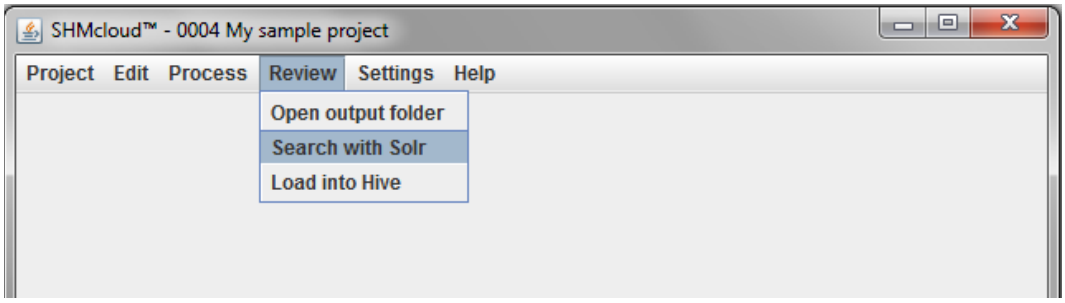
Step 1 - Stage your project.



Step 2 - Process locally, and wait for your project to complete processing.



Step 3 - Review your output in Solr.



➡ When processing is complete all of your output will be usable.

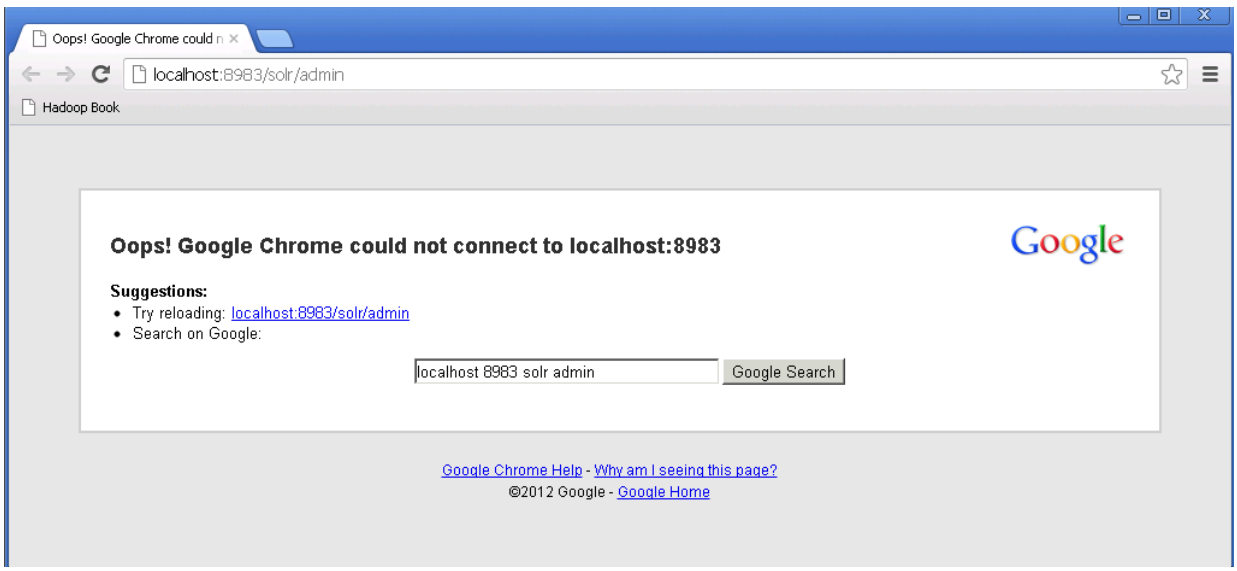
➡ Click Review. All of output from your project has gone into your output folder as it normally does. Additionally, if everything was set up properly, then all of your output was also sent to apache-solr-3.6.1, in the form of searchable text.

➡ Click “Search with Solr”.

➡ Note: you may also go to http://localhost:8983/solr/admin/ to get to the same Solr search screen.

Error Messages while attempting to access Solr

While you are attempting to open your Solr search window, you may encounter a message telling you that your browser could not connect to “localhost:8983”. This means that your solr machine is currently turned off, or you might not even have Solr installed on your computer..

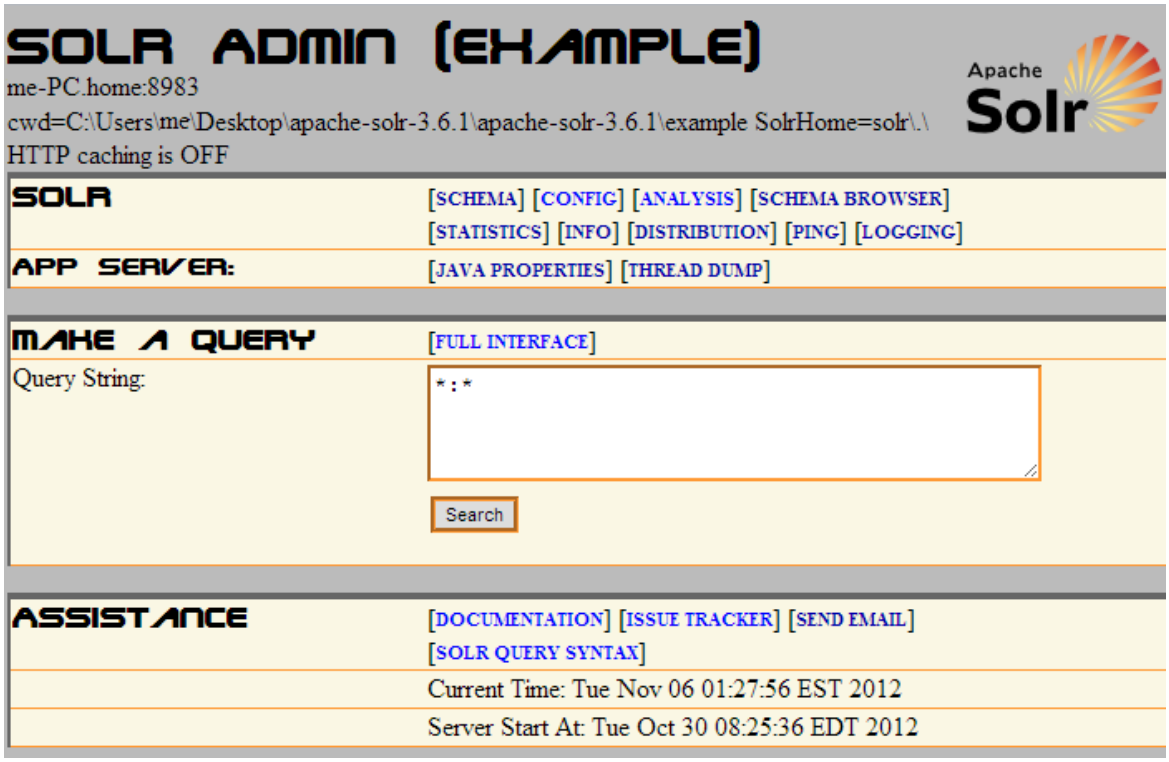


You can easily fix this by following the steps outlined above for turning on your Solr Search Server. Of course if you have not yet installed Solr on your machine, then you will need to follow those steps as well.

➡ If there the error 404 comes up, don't worry about it. Simply follow the link to the proper page.

Error 404 - Not Found.
No context on this server matched or handled this request.
Contexts known to this server are:
/solr ---> org.mortbay.jetty.webapp.WebAppContext@db7b35{/solr,file:/C:/Users/.me/Desktop/apache-solr-3.6.1/apache-solr-3.6.1/example/webapps/solr.war}

The Apache-Solr screen will open, as seen below.



15.3 How to search through your output in the Solr Search Server:

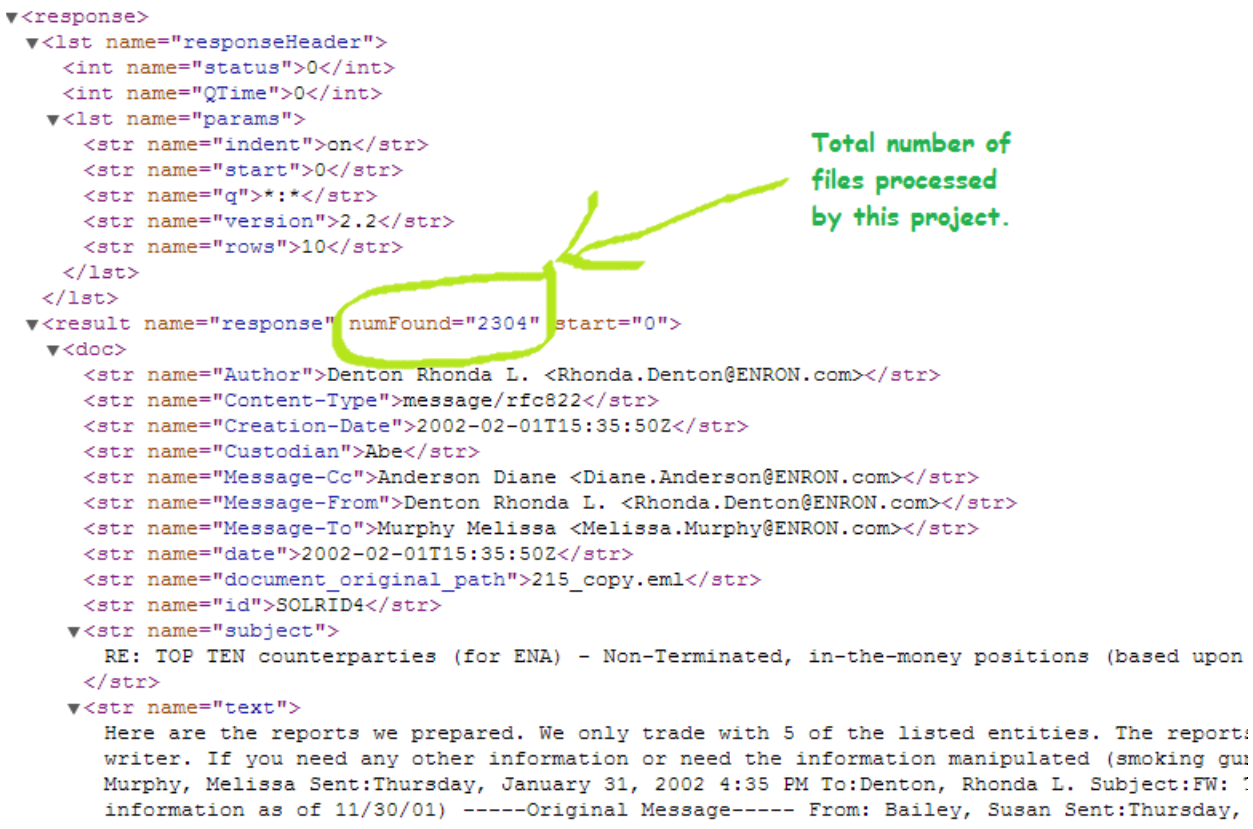
Viewing all of your processed documents at one time:

➡ Searches are done by entering a string in the “Make A Query / Query String” box. If you do not configure your search for anything specific, you should be able to see everything that was processed, since ***:*** is a search for everything.

➡ While all documents are passed to Solr, the default documents per page is 10.

➡ To see the actual number of documents that you processed, query *:* (all of them).

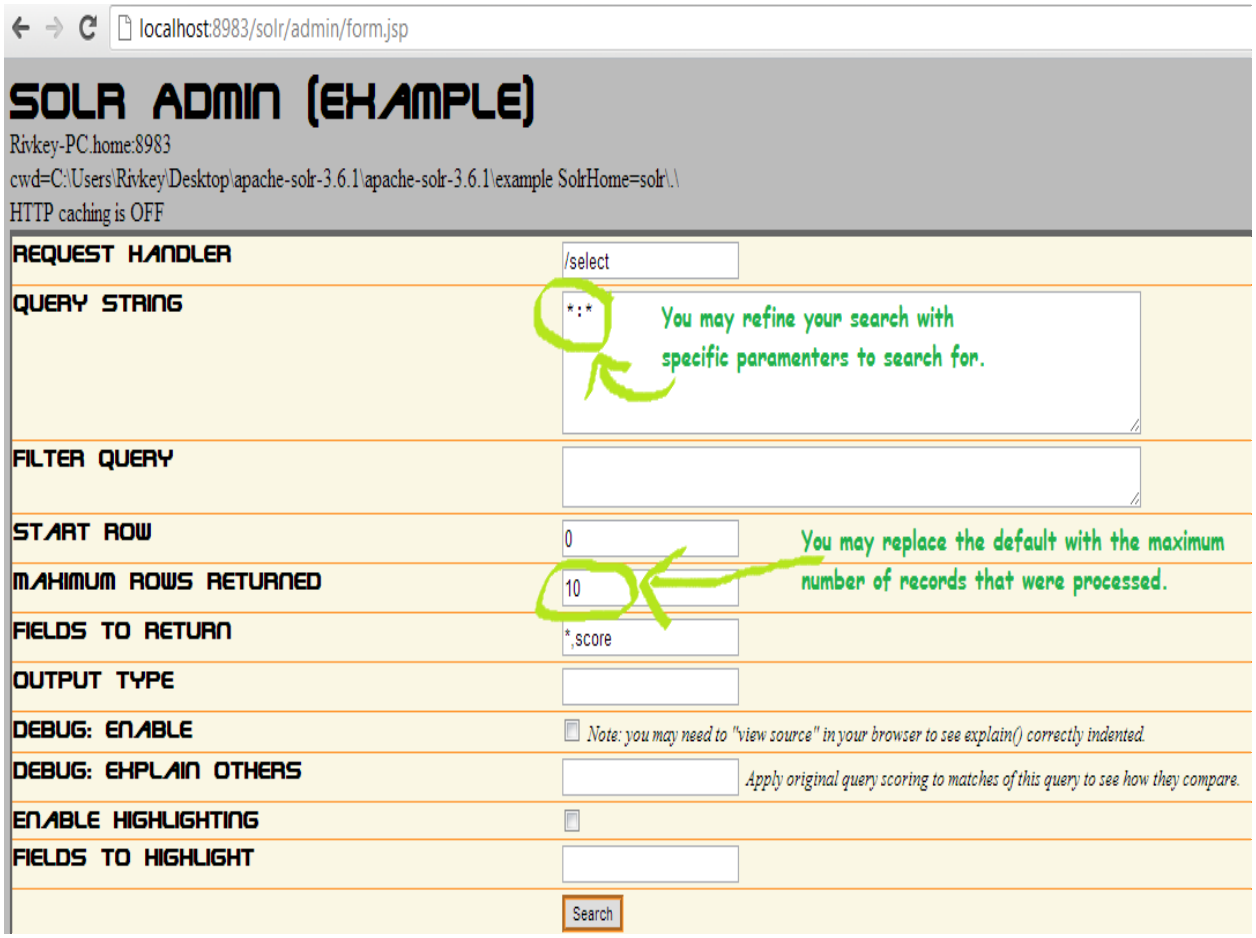
➡ Use your back-arrow to go back to the Query screen.




```
▼<response>
  ▼<lst name="responseHeader">
     <int name="status">0</int>
     <int name="QTime">0</int>
    ▼<lst name="params">
       <str name="indent">on</str>
       <str name="start">0</str>
       <str name="q">*:*</str>
       <str name="version">2.2</str>
       <str name="rows">10</str>
     </lst>
   </lst>
  ▼<result name="response" numFound="2304" start="0">
    ▼<doc>
       <str name="Author">Denton Rhonda L. <Rhonda.Denton@ENRON.com></str>
       <str name="Content-Type">message/rfc822</str>
       <str name="Creation-Date">2002-02-01T15:35:50Z</str>
       <str name="Custodian">Abe</str>
       <str name="Message-Cc">Anderson Diane <Diane.Anderson@ENRON.com></str>
       <str name="Message-From">Denton Rhonda L. <Rhonda.Denton@ENRON.com></str>
       <str name="Message-To">Murphy Melissa <Melissa.Murphy@ENRON.com></str>
       <str name="date">2002-02-01T15:35:50Z</str>
       <str name="document_original_path">215_copy.eml</str>
       <str name="id">SOLRID4</str>
      ▼<str name="subject">
         RE: TOP TEN counterparties (for ENA) - Non-Terminated, in-the-money positions (based upon
       </str>
      ▼<str name="text">
         Here are the reports we prepared. We only trade with 5 of the listed entities. The reports
         writer. If you need any other information or need the information manipulated (smoking gu
         Murphy, Melissa Sent:Thursday, January 31, 2002 4:35 PM To:Denton, Rhonda L. Subject:FW:
         information as of 11/30/01) -----Original Message----- From: Bailey, Susan Sent:Thursday,
```

➡ In the example above, 2304 files were processed by this project run. "numfound” holds the value for the number of files that were processed.

➡ While the documents per page are fixed to 10, this can be changed by clicking on “Full interface” in the “Make a Query” section, and then **Maximum Rows Returned.**



➡ As shown in the examples above, if you do your search with *:* in the Query String box and 10 in the “Maximum Rows Returned” box, with “Start Row” set at 0, then your returned results will simply be the first 10 files that were processed by the SHMcloud™ Player.

➡ If you change the “10” in the “Maximum Rows Returned” box to a higher number, then it will increase the amount of Records available for you to review at one time. So in our example above, changing “10” to “2304” will return 2304 files. Or if your search is refined, then it will search through 2304 records at once.

In a nutshell, you can adjust the starting point for your search by changing the 0 in the Start Row, and you can adjust the number of records / files to review by changing the Maximum Rows Returned. You can also refine your search by changing the *:*, which we will be discussing shortly.

Searching all of your data manually, while using a standard search function:

If you wish to view all 2304 records at once for the purpose of scrolling through the output, you can do so by simply changing the “10” to “2304”, and pushing the “Search” button.

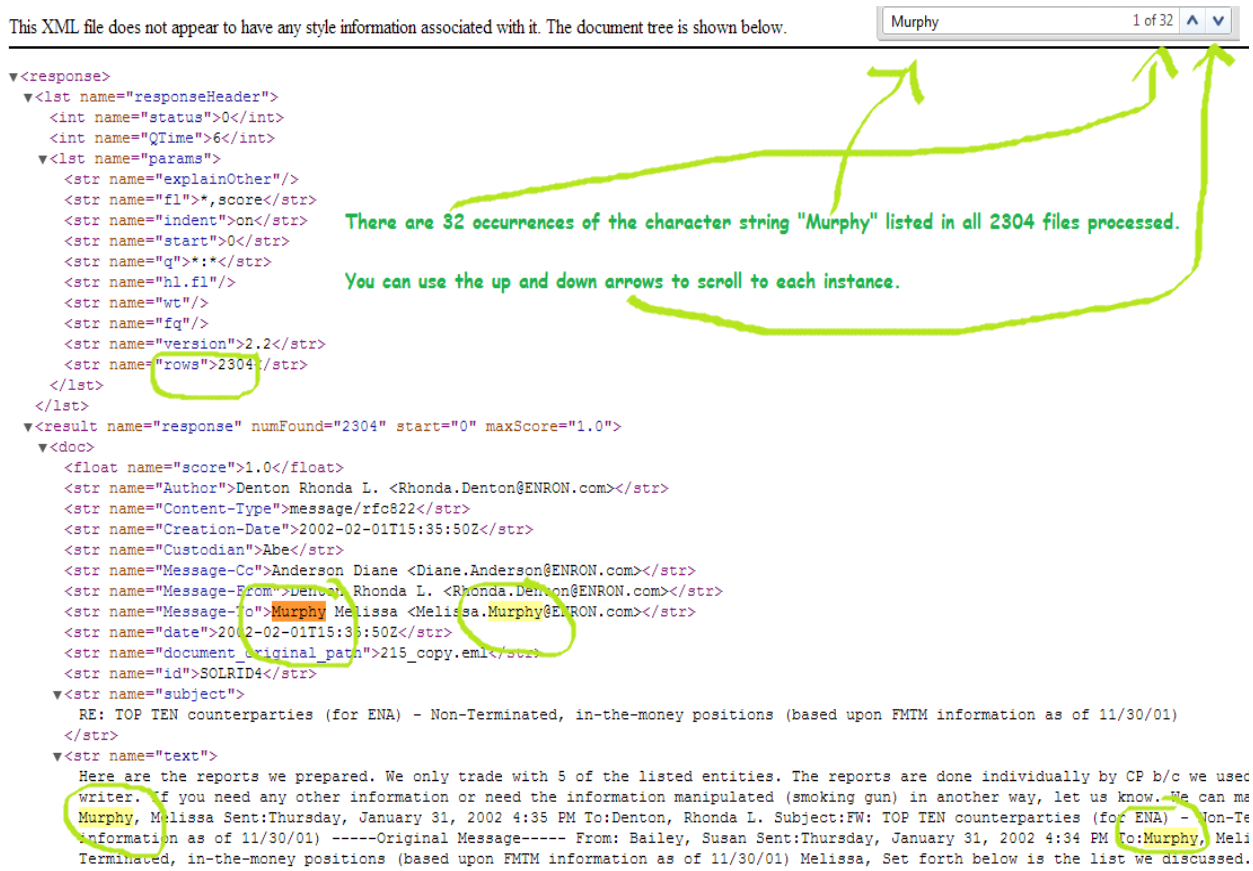
Note: Standard Search Function available on all computers

CTRL + F in Windows or COMMAND + F on a Mac....

This should open the Find/Search function in Solr.

Then type your query text into the search box in the upper right corner of your screen. You will then be able to see all of the occurrences of that text within your Solr output.

Using the output from our sample project, the following example shows how to use the standard search function to pinpoint specific character strings.

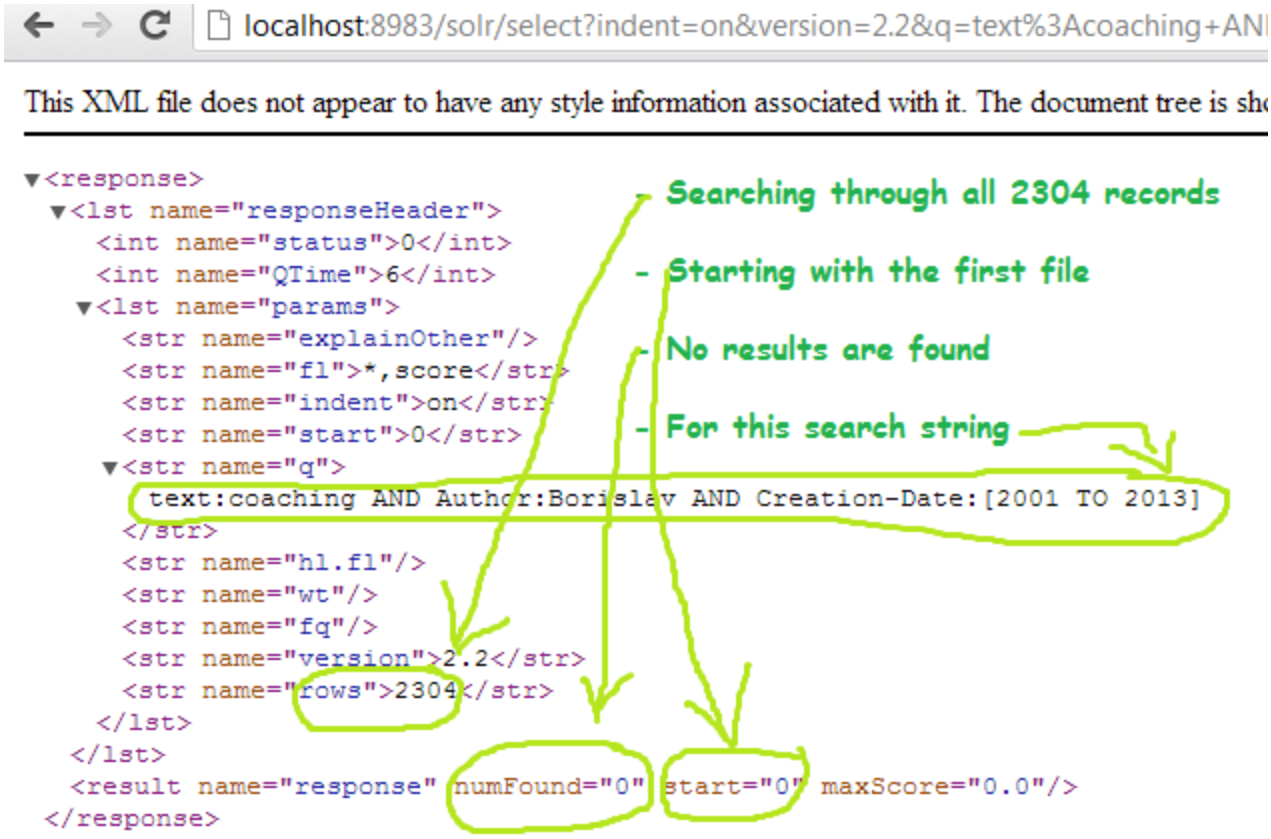

This XML file does not appear to have any style information associated with it. The document tree is shown below.
Murphy
1 of 32
There are 32 occurrences of the character string "Murphy" listed in all 2304 files processed.
You can use the up and down arrows to scroll to each instance.

Refining your search, using the Solr Search features:

For the exclusive purpose of understanding formatting, here is a sample search string to be placed in the Query String box in place of the “everything” search that *:* does.

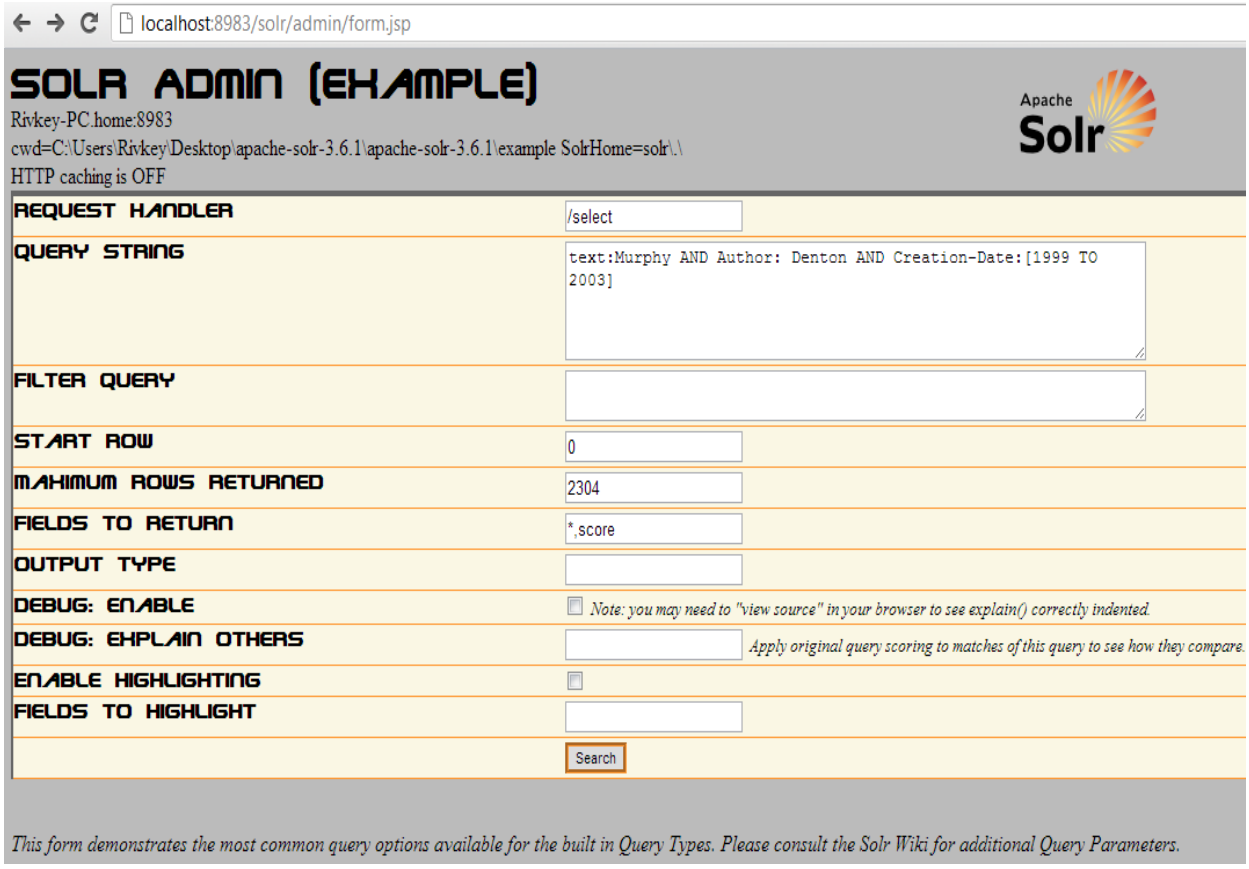
text:coaching AND Author:Borislav AND Creation-Date:[2001 TO 2013]

Since none of the parameters in this example exist in our current dataset, the results will return negative, as seen in the following example:



If Solr returns a negative result, as seen above, then you might want to broaden your search parameters by putting in fewer restrictions.

Now let's test examples to refine our search using our current sample output.

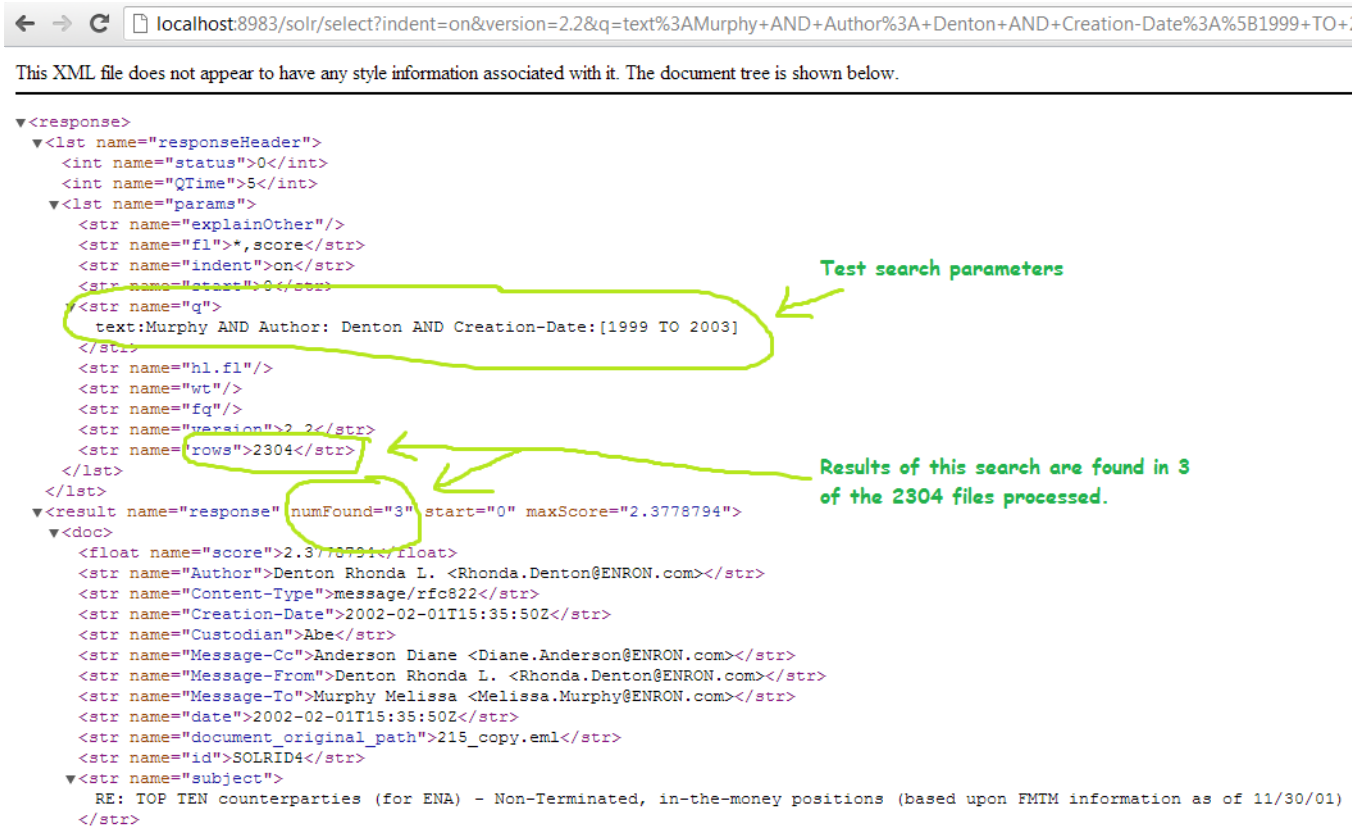

localhost:8983/solr/admin/form.jsp
SOLR ADMIN (EXAMPLE)
Rivkey-PC.home:8983
cwd=C:\Users\Rivkey\Desktop\apache-solr-3.6.1\apache-solr-3.6.1\example SolrHome=solr\.\
HTTP caching is OFF
Apache Solr
REQUEST HANDLER
/select
QUERY STRING
text:Murphy AND Author: Denton AND Creation-Date:[1999 TO 2003]
FILTER QUERY
START ROW
0
MAHIMUM ROWS RETURNED
2304
FIELDS TO RETURN
*,score
OUTPUT TYPE
DEBUG: ENABLE
Note: you may need to "view source" in your browser to see explain() correctly indented.
DEBUG: EHPLAIN OTHERS
Apply original query scoring to matches of this query to see how they compare.
ENABLE HIGHLIGHTING
FIELDS TO HIGHLIGHT
Search
This form demonstrates the most common query options available for the built in Query Types. Please consult the Solr Wiki for additional Query Parameters.

The example above tests all 2304 files and uses the query string:

text:Murphy AND Author: Denton AND Creation-Date:[1999 TO 2003]

Our output results should list for us all of the files that contain "Murphy", where the Author is "Denton", and were created in the date range of 1999 to 2003.

Clicking the Search button at the bottom of the screen should return the results seen in the image below.

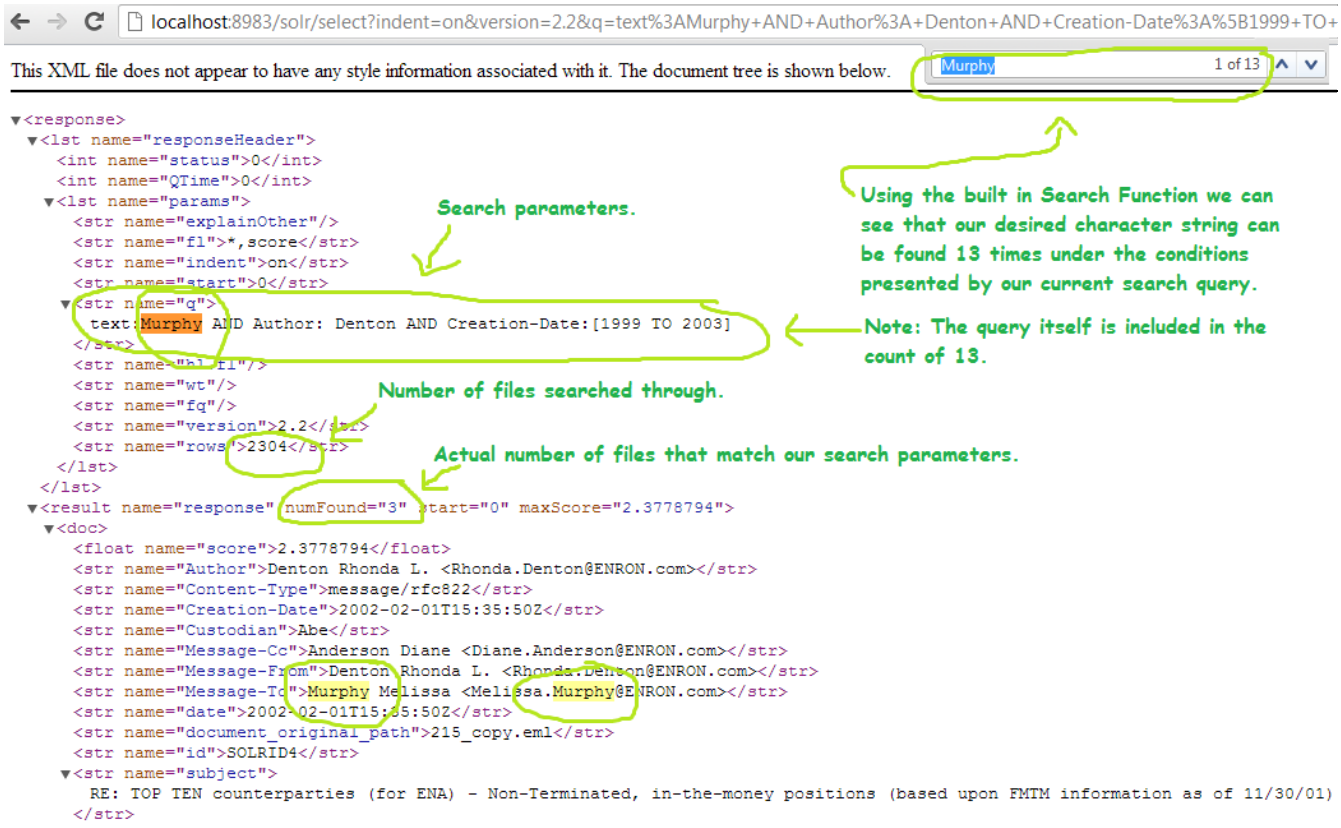


As you can see from the example above, when searching through all 2304 files for “Murphy”, in documents that were created by “Denton”, during the years 1999 to 2003, the output shows that the results of this search can be found in 3 records.

However, it is still difficult to find our desired results.

Taking advantage of the manual built-in Search Function, as we discussed above, I have chosen to further refine my search on Murphy.

Below you can see the results of using **CTRL + F in Windows** or **COMMAND + F** on a Mac,


localhost:8983/solr/select?indent=on&version=2.2&q=text%3AMurphy+AND+Author%3A+Denton+AND+Creation-Date%3A%5B1999+TO+
This XML file does not appear to have any style information associated with it. The document tree is shown below.
Murphy
1 of 13
Using the built in Search Function we can see that our desired character string can be found 13 times under the conditions presented by our current search query.
Search parameters.
Note: The query itself is included in the count of 13.
Number of files searched through.
Actual number of files that match our search parameters.

```
▼<response>
 ▼<lst name="responseHeader">
    <int name="status">0</int>
    <int name="QTime">0</int>
   ▼<lst name="params">
      <str name="explainOther"/>
      <str name="fl">*,score</str>
      <str name="indent">on</str>
      <str name="start">0</str>
     ▼<str name="q">
       text:Murphy AND Author: Denton AND Creation-Date:[1999 TO 2003]
      </str>
      <str name="hl.fl"/>
      <str name="wt"/>
      <str name="fq"/>
      <str name="version">2.2</str>
      <str name="rows">2304</str>
    </lst>
  </lst>
 ▼<result name="response" numFound="3" start="0" maxScore="2.3778794">
  ▼<doc>
     <float name="score">2.3778794</float>
     <str name="Author">Denton Rhonda L. <Rhonda.Denton@ENRON.com></str>
     <str name="Content-Type">message/rfc822</str>
     <str name="Creation-Date">2002-02-01T15:35:50Z</str>
     <str name="Custodian">Abe</str>
     <str name="Message-Cc">Anderson Diane <Diane.Anderson@ENRON.com></str>
     <str name="Message-From">Denton Rhonda L. <Rhonda.Denton@ENRON.com></str>
     <str name="Message-To">Murphy Melissa <Melissa.Murphy@ENRON.com></str>
     <str name="date">2002-02-01T15:35:50Z</str>
     <str name="document_original_path">215_copy.eml</str>
     <str name="id">SOLRID4</str>
    ▼<str name="subject">
       RE: TOP TEN counterparties (for ENA) - Non-Terminated, in-the-money positions (based upon FMTM information as of 11/30/01)
     </str>
```

Using the built-in search CTRL + F, I typed “Murphy” into the search box. We can now see that Murphy appears 13 times in our search for Murphy under the conditions presented in the search string.

Please also note that the count of 13 also includes the appearance of “Murphy” in our original query string.

Additional Search Options in Apache-Solr:

Now that we understand the basics of searching and refining our searches, here are a few other examples that you might want to use as you search through records using Apache-Solr in conjunction with your SHMcloud™ Player.

Your Query String can be refined to search for items with only one condition, or many conditions. Be advised that the more conditions you place on your search, the fewer output results you will have. If your search renders zero results, then you will need to broaden your search.

Here are some examples of search options to place in your Query String box of your Solr search window. The results recorded here are based on the Sample Project *sample_freeeed_windows.project,* that we recently processed.

Custodian Search:

Custodian:Abe when you process this search using all 2304 of our sample output, the results should show **numFound="4".** This shows us that Abe is the Custodian of 4 of the files that were processed by this project.

Custodian:Jackie will return **numFound="2178"**, showing us that Jackie appears as the custodian 2178 times in our dataset of 2304 files.

Date Range Search:

In our query, **Creation-Date:[2001 TO 2002]** may return the result **numFound="435"**.

You may be wondering why only 435 records were returned with this search. The answer is simple. (1) Only 435 records have that date range associated with it. The date might be found in the creation of an email, or possibly in the creation of a document within the dataset. (2) If you or anyone else re-saved any of the documents being processed, the date associated with that file may very well be affected, which can alter the results of your search. (3) Additionally, if your file somehow has no dates associated with it, then it will be excluded from this search altogether.

Test String Search:

Doing a search on money, **text:Money** returned a count of 53 files, **numFound="53"**. I further enhanced this search by using the built-in search function **CTRL+F**, and I found that in those 53 files, “money” appears 158 times. Using the up and down arrows associated with that search, I am able to scroll to all locations where “money” is referenced.

Author Search:

Searching for Denton as the Author or my output **Author: Denton** has returned 5 records **numFound="5"**. Enhancing my search with **CTRL+F**, I have found that Denton is mentioned 26 times (including my original search string), in the query output.

Combining Search Options:

Your search options can be combined for a single query search, or for separate individual searches.

Such examples could be:

text:Murphy AND Author: Denton AND Creation-Date:[1999 TO 2001]

↳ Search for all documents that contain “Murphy”, that were created by“Denton”, and range between 1999 and 2001.

Or, if you wish to broaden that search to include more possibilities that will not be limited by a creation date:

text:Murphy AND Author: Denton

↳ Search for all documents that contain “Murphy” and were created by“Denton”.

As you can see from our examples above, you are able to narrow down or broaden your search options using the Apache-Solr Search Server, in conjunction with the SHMcloud™ Player.

Licensing

Copyright 2012, SHMsoft, Inc.

END USER SOFTWARE LICENSE

IMPORTANT READ BEFORE INSTALLING OR OPERATING THIS PRODUCT

LICENSEE AGREES TO BE BOUND BY THE TERMS OF THIS AGREEMENT BY INSTALLING, HAVING INSTALLED, COPYING, OR OTHERWISE USING THE PRODUCT. IF LICENSEE DOES NOT AGREE, DO NOT INSTALL OR USE THE PRODUCT.

1. Scope. This License applies to the software product (Software) you have licensed from SHMsoft, Inc. (SHMsoft). The Software is licensed for use in conjunction with SHMSOFT hardware which together with the Software will be referenced as the Product. This License is a legal agreement between SHMSOFT and the single entity (Licensee) that has acquired the Software from SHMSOFT under these terms and conditions. The Software incorporates certain third party software programs subject to the terms and restrictions of the applicable licenses identified herein.

2. License Grant. Subject to the terms of this License, SHMSOFT grants to Licensee a perpetual, non-exclusive, non-transferable license to use the Software for which Licensee has paid the required license fees in object code form for Licensee's internal business purposes. Other than as specifically described herein, no right or license is granted to any of SHMSOFT's trademarks, patents, copyrights, or other intellectual property rights and SHMSOFT retains all rights not granted herein. The Software incorporates certain third party open source software. The protections given to SHMSOFT under this License also apply to the suppliers of this third party software.

3. Restrictions.

(a) The Software, documentation and the associated copyrights and other intellectual property rights are owned by SHMSOFT or its licensors and are protected by law and international treaties. Licensee may not copy or translate the documentation provided with the Software or available online at http://www.shmsoft.com (Documentation) without SHMSOFT's prior, written consent. Licensee may install, use, access, display and run the Software only in the manner in which it has been licensed, including but not limited to any restrictions on number of protected applications, number or type of licensed devices, number of users, bandwidth, non-production use or database restrictions. SHMSOFT reserves the right to audit Licensee's use of the

Software or authorize others to conduct such an audit on its behalf and to disable any application or functionality that has not been specifically licensed.

(b) Certain portions of the Software include third party software modules as identified in the applicable Software release notes, including but not limited to, Apache License, Version 2.0 found at http://www.apache.org/licenses/LICENSE-2.0 and MySQL licensed from MySQL AB and JavaTM licensed from Sun Microsystems, and are subject to additional limitations imposed by those third parties (Third Party Software). You may not use these files except in compliance with the Licenses. Unless required by applicable law or agreed to in writing, software distributed under the Apache 2.0 License is distributed on an "AS IS" BASIS, WITHOUT WARRANTIES OR CONDITIONS OF ANY KIND, either express or implied. See the License for the specific language governing permissions and limitations under the License. Certain portions of the Software may also include geographical or other data (Data). Licensee agrees that it will only use such Third Party Software or Data in conjunction with the Product and not as standalone software. Licensee will not (i) copy the Third Party Software or Data onto any public or distributed network; (ii) use the Third Party Software or Data separately to operate in or as a time-sharing, outsourcing, service bureau, application service provider or managed service provider environment; (iii) use the Third Party Software or Data as a general server, as a standalone application or with applications other than the Software under this license; (iv) change any proprietary rights notices which appear in the Third Party Software or Data; or (v) modify the Third Party Software or Data.

(c) Licensee may not copy (except to make one archival copy for backup and disaster recover purposes), modify, sell, sub-license, rent or transfer the Software, Data or any associated Documentation to any third party. Licensee may not disassemble, reverse compile or reverse engineer the Software or any Data incorporated in the Software or encourage others to do so except as required by law for interoperability purposes, and then only after Licensee has given Supplier an opportunity to provide information or software necessary to resolve such interoperability issues.

4. Export Control. SHMSOFT's standard Product incorporates cryptographic software. Licensee agrees to comply with the Export Administration Act, the Export Control Act, all regulations promulgated under such Acts, and all other US government regulations relating to the export of technical data and equipment and products produced therefrom which are applicable to Licensee. In countries other than the US, Licensee agrees to comply with the local regulations regarding importing, exporting or using cryptographic software.

5. This Software is provided AS-IS, WITHOUT WARRANTIES OR CONDITIONS OF ANY KIND, EITHER EXPRESS OR IMPLIED, INCLUDING, WITHOUT LIMITATION, ANY WARRANTIES OR CONDITIONS OF TITLE, NON-INFRINGEMENT, MERCHANTABILITY OR FITNESS FOR A PARTICULAR PURPOSE. NEITHER THE LICENSOR NOR ITS SUPPLIERS WILL BE LIABLE TO THE FOUNDATION OR ITS LICENSEES FOR ANY DIRECT, INDIRECT, INCIDENTAL, SPECIAL, EXEMPLARY, OR CONSEQUENTIAL DAMAGES (INCLUDING WITHOUT

LIMITATION LOST PROFITS), HOWEVER CAUSED AND ON ANY THEORY OF LIABILITY, WHETHER IN CONTRACT, STRICT LIABILITY, OR TORT (INCLUDING NEGLIGENCE OR OTHERWISE) ARISING IN ANY WAY OUT OF THE USE OR DISTRIBUTION OF THE WORK OR THE EXERCISE OF ANY RIGHTS GRANTED HEREUNDER, EVEN IF ADVISED OF THE POSSIBILITY OF SUCH DAMAGES.

9. Non-Production Use Software. If Licensee purchases an SHMSOFT Product or licenses SHMSOFT Software designated as non-production, non-commercial, lab or development Product in the applicable purchase order, quote or the license file for such Product or Software, Licensee may use the Software included with such Product to conduct testing and development in Licensee's non-production environment only and not to manage data traffic or applications in the ordinary course of Licensee's business.

10. Evaluation Software. If the Software is Evaluation Software, notwithstanding any other terms to the contrary in this Agreement, Licensee may use the Software only for its internal demonstration, test or evaluation purposes and not in a production environment. Notwithstanding any terms to the contrary in this License, Evaluation Software is provided on an AS IS basis and has a non-perpetual time limited license that will time-out and disable the Software upon expiration of the evaluation period.

11. Termination. The license granted in Section 2 is effective until terminated and will automatically terminate if Licensee fails to comply with any of the terms and conditions set forth herein. Upon termination, Licensee will destroy the Software and documentation and all copies or portions thereof.

12. Support. Maintenance and support of the Software is not provided under this License and must be purchased separately subject to SHMSOFT's support policies available at http://www.SHMsoft.com Where Licensee has purchased maintenance and support for a Product, the term Software under this License will include any published updates, corrections, new releases and new versions of such Software (collectively Updates), provided that Licensee is otherwise entitled to access and use such Updates pursuant to the applicable maintenance and support contract. Licensee may only use the Updates on Products for which Licensee is the original end user or other Products which include Software to which Licensee holds a valid license, and only on equipment for which Licensee has purchased maintenance and support.

13. Miscellaneous. This License will be governed by the laws of the State of Texas, USA without regard to its choice of law rules. The provisions of the U.N. Convention for the International Sale of Goods and the Uniform Computer Information Transactions Act, in whatever form adopted, will not apply and the parties specifically opt out of the application of such laws. In the event of any dispute arising out of or relating to this Agreement, the parties shall seek to settle the dispute via direct discussions. If a dispute cannot be settled through direct discussions, the parties agree to first endeavor to settle the dispute via voluntary nonbinding mediation, before resorting to arbitration. A mediator will be selected by voluntary agreement of both parties, or in the event

both parties cannot agree on a mediator, a mediator will be selected in accordance with the rules of JAMS. The mediation shall be held in Houston, Texas. Each party shall bear its own costs and expenses and an equal share of the administrative and other fees associated with the mediation. Any dispute that remains unresolved following mediation shall be settled by arbitration administered by the JAMS in accordance with its Comprehensive Arbitration Rules. The place of arbitration shall be Houston, Texas. Judgment upon the award rendered by the arbitrator(s) may be entered in any court having jurisdiction thereof. The arbitrator(s) shall award to the prevailing party, if any, as determined by the arbitrator(s), all of its costs and fees. "Costs and fees" mean all reasonable pre-award expenses of the arbitration, including the arbitrators' fees, administrative fees, travel expenses, out-of-pocket expenses such as copying and telephone, court costs, witness fees, and attorneys' fees. In rendering the award, the arbitrator(s) shall determine the rights and obligations of the parties according to the substantive and procedural laws of the State of Texas. The foregoing alternative dispute resolution provisions will not apply to claims or actions related to the infringement, misappropriation or violation of SHMSOFTs intellectual property rights or those of its third party licensors and such actions may be brought in any court of competent jurisdiction. Any provisions found to be unenforceable will not affect the enforceability of the other provisions contained herein, but will instead be replaced with a provision as similar in meaning to the original as possible. This License constitutes the entire agreement between the parties with regard to its subject matter. No modification will be binding unless in writing and signed by the parties.

14. Acknowledgements. The Software includes Data and software developed by third parties subject to separate licenses. Please refer to the Acknowledgement section found in the Software Documentation available at http://SHMsoft.com.

15. GPL. Limited portions of the software contain software code subject to the GNU GPL Version 2 available at http://www.gnu.org/licenses/gpl.html. Please refer to the Acknowledgement section found in the Software documentation for the specific references. GPL software is not subject to the restrictions set forth in this License but is licensed separately under the GPL. Only those portions of the software that are licensed under the GPL are subject to the GPL license. All other software code is subject to the restrictions set forth elsewhere in this License. Furthermore, those portions of the software that are licensed under the GPL are subject to the remaining terms and conditions of the License to the extent that those terms are not inconsistent with the terms of the GPL.